

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:32 : Search time 234.98 Seconds
(without alignments)
250.084 Million cell updates/sec

Title: US-08-956-518A-103
Sequence: 1 CAGGCGCCGACATGCTGCC.....CAGAGTCTCTGCTACATCG 641

Scoring table: IDENTITY_NUC
Attached: 176461 segs, 45838279 residues

base: Issued Patents, NA: *
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/lna/5C.COMB.seq.*
4: /cgn2_6/ptodata/1/lna/5D.COMB.seq.*
5: /cgn2_6/ptodata/1/lna/PCUS9.COMB.seq.*
6: /cgn2_6/ptodata/1/lna/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	108	16.8	1876	4	US-08-466-589-7	Sequence 7, Appl1
2	40.2	6.3	2990	3	US-08-572-851-1	Sequence 1, Appl1
3	36.6	5.7	1915	3	US-08-466-855A-5	Sequence 5, Appl1
4	36.6	5.7	1915	4	US-08-466-589-11	Sequence 11, Appl1
5	35.4	5.5	1942	4	US-08-604-989A-11	Sequence 11, Appl1
6	35.2	5.5	31571	2	US-08-323-443B-1	Sequence 1, Appl1
7	34.6	5.4	2374	4	US-08-466-589-5	Sequence 5, Appl1
8	34	5.3	3223	4	US-07-980-528-1	Sequence 1, Appl1
9	33.8	5.3	1018	1	US-08-444-083-6	Sequence 6, Appl1
10	33.8	5.3	1018	1	US-08-286-304-6	Sequence 6, Appl1
11	33.8	5.3	1018	1	US-08-442-745-6	Sequence 6, Appl1
12	33.8	5.3	1018	2	US-08-443-129-6	Sequence 6, Appl1
13	33.8	5.3	1018	2	US-08-443-952-6	Sequence 6, Appl1
14	33.8	5.3	1018	2	US-08-443-130-6	Sequence 6, Appl1
15	33.8	5.3	43280	4	US-08-804-227C-1	Sequence 1, Appl1
16	33.8	5.3	1018	5	PCT-US95-04467-6	Sequence 6, Appl1
17	33.6	5.2	8438	1	US-07-945-283-1	Sequence 1, Appl1
18	33.6	5.2	1597	1	US-08-166-316-1	Sequence 1, Appl1
19	33.4	5.2	2064	2	US-08-343-428-1	Sequence 1, Appl1
20	33.4	5.2	1152	5	PCT-US93-06251-51	Sequence 1, Appl1
21	33.2	5.2	388	1	US-07-626-618A-1	Sequence 1, Appl1
22	33.2	5.2	388	1	US-08-333-977-1	Sequence 1, Appl1
23	33.2	5.2	2638	3	US-08-306-691B-46	Sequence 1, Appl1
24	33.2	5.1	11219	3	US-07-642-734C-1	Sequence 1, Appl1
25	33	5.1	11558	5	PCT-US93-06251-23	Sequence 23, Appl1
26	32.8	5.1	1161	3	US-08-466-855A-1	Sequence 1, Appl1
27	32.6	5.1	1161	3	US-08-086-439C-2	Sequence 2, Appl1
28	32.6	5.1	12001	3	US-08-434-877-2	Sequence 11, Appl1
29	32.6	5.1	4257	4	US-08-458-568A-11	Sequence 1, Appl1
30	32.6	5.1	2923	2	US-08-690-473-1	Sequence 1, Appl1
31	32.4	5.1	2923	4	US-07-969-847-6	Sequence 6, Appl1
32	32.4	5.1	2923	4	US-07-969-847-7	Sequence 7, Appl1
33	32.2	5.0	11613	1	US-08-484-044-10	Sequence 10, Appl1
34	32.2	5.0	1864	1	US-08-353-948-1	Sequence 2, Appl1
35	32	5.0	6961	3	US-08-727-034-2	Sequence 1, Appl1
36	32	5.0	1578	3	US-08-681-129-1	Sequence 1, Appl1
37	32	5.0	3836	4	US-08-216-260-1	Sequence 1, Appl1

ALIGNMENTS

38	31.8	5.0	2505	1	US-07-977-434-9	Sequence 9, Appl1
39	31.8	5.0	2504	1	US-08-073-384C-3	Sequence 3, Appl1
40	31.8	5.0	2504	1	US-08-254-359A-3	Sequence 3, Appl1
41	31.8	5.0	2640	1	US-08-384-490-30	Sequence 30, Appl1
42	31.8	5.0	2504	2	US-08-483-043-3	Sequence 3, Appl1
43	31.8	5.0	2640	3	US-08-459-383-30	Sequence 30, Appl1
44	31.8	5.0	2505	3	US-08-458-819-9	Sequence 9, Appl1
45	31.8	5.0	2504	3	US-08-481-238-3	Sequence 3, Appl1

RESULT 1
US-08-466-589-7
Sequence 7, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSP Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 73..1581
US-08-466-589-7
Query Match 16.8%; Score 108; DB 4; Length 1876;
Best Local Similarity 95.7%; Pred. No. 7.6e-22;
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 526 CTCGAGATATGCGATGAGCGCTTTGAGCGCACATTCACACATGATGCTGATTC 585
DB 414 CTATACAGCGCTGATGAGCGCTTTGAGCGCACATTCACACATGATGCTGATTC 473
QY 586 TTCTGGGCAATGGCAGTACCTGCTCCAGGACATATTCAGAGTTCCTGCTACATCG 641

Db 474 TTCTGGGCGATTCAGGAGTCTGCTCCAGGATATCTACAGAGTCTCTGCTACATCG 529

RESULT 2
US-08-572-951-1

Sequence 1, Application US/08572951
Patent No. 5824790

GENERAL INFORMATION:

APPLICANT: KEELING, PETER L.

APPLICANT: KNIGHT, MARY E.

APPLICANT: GUAN, HANPING

TITLE OF INVENTION: MODIFICATION OF STARCH

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN

ADDRESSEE: Intellectual Property Group of

ADDRESSEE: Pillsbury Madison & Sutro LLP

STREET: 1100 New York Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/572,951

FILING DATE: 15-DEC-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/346,602

FILING DATE: 29-NOV-1994

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/263,921

FILING DATE: 21-JUN-1994

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Paul N. Kokulis

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 222957/1.02.15C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2990 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-572-951-1

Query Match
Best Local Similarity 57.6%; Score 40.2; DB 3; Length 2990;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 58 GCTCCACGAGGCTACAGGGGGGGGAGAGAGTGCAGAGCTCGGGGGGGGCC 117
DB 693 GCTCCACGAGGCTACAGGGGGGGGAGAGAGTGCAGAGCTCGGGGGGGGCC 752
QY 118 CGGCTGATGAGCGCGGCGCATGACAGCGGCGTGGGACTGGCTCTTTCGGCGCCCTTC 177
DB 753 GCGCGCCCTGCGACCGCGGCTGCTGCGCGCCCGCCCTGCTGCGCGCCCTTCCTCGGCGCC 812
QY 178 GCGCG 182
DB 813 GCGCG 817

RESULT 3
US-08-496-855A-5
Sequence 5, Application US/08496855A
Patent No. 5801232

GENERAL INFORMATION:

APPLICANT: Elliot, Kathryn J.

APPLICANT: Ellis, Steven B.

APPLICANT: Harpold, Michael M.

TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE

RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brown, Martin, Haller & McClain

ADDRESSEE: 1660 Union Street

STREET: San Diego

CITY: San Diego

STATE: CA

COUNTRY: U.S.A.

ZIP: 92101-2926

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/496,855A

FILING DATE: 20-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/149,503

FILING DATE: 08-NOV-1993

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/028,031

FILING DATE: 08-MAR-1993

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-9369B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1915 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 87..1583

Query Match
Best Local Similarity 57.4%; Score 36.6; DB 3; Length 1915;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 527 TGGGATATGCTATGAGGCGCTTACGCGCATTCACACTACAGTGTGGATTC 566
DB 438 TACACACAGCGCGGACGACCTATGAGGTGTCTTACACCACTTGATAGTCGCTC 497
QY 587 TCTGGGCAATTCGACGACCTCCAGGCAATTCAGAGTCTGCTACATCG 641
DB 498 AACGCGAGGCTGCTGTGGCTGCGCCCTGCGCATCTACAGGCGCTGCAAGATG 552
RESULT 4
US-08-466-589-11
Sequence 11, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM.nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:13 ; Search time 234.98 Seconds
(Without alignments)
39.015 Million cell updates/sec

Title: US-08-956-518a-99

Perfect score: 100
Sequence: 1 CTCCTTCTACTGCTGATGAG.....CTGCTCCAGGTAAGCTGCA 100

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

abase :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCUS9.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Dackille1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83.4	83.4	1876	4	US-08-466-589-7	Sequence 7, App1
2	24.2	24.2	1384	1	US-07-607-538C-1	Sequence 1, App1
3	23.6	23.6	11601	4	US-08-222-617A-3	Sequence 3, App1
4	23.6	23.6	11601	4	US-08-222-617A-24	Sequence 24, App1
5	23	23.0	46899	3	US-08-471-119A-1	Sequence 1, App1
6	22.8	22.8	1810	1	US-08-299-849B-20	Sequence 20, App1
7	22.4	22.4	35100	4	US-08-770-379-17	Sequence 17, App1
8	22.2	22.2	4566	3	US-08-484-101B-37	Sequence 37, App1
9	22	22.0	1447	1	US-08-236-427-2	Sequence 2, App1
10	22	22.0	1444	1	US-08-236-427-9	Sequence 9, App1
11	21.8	21.8	7432	1	US-08-260-1	Sequence 1, App1
12	21.8	21.8	1400	1	US-08-464-164-1	Sequence 1, App1
13	21.8	21.8	4046	3	US-08-706-936-1	Sequence 1, App1
14	21.8	21.8	1400	3	US-08-338-057-1	Sequence 1, App1
15	21.8	21.8	7432	4	US-08-461-503-1	Sequence 1, App1
16	21.8	21.8	1400	4	US-08-668-416-1	Sequence 1, App1
17	21.8	21.8	2042	4	US-08-616-844-5	Sequence 5, App1
18	21.8	21.8	2042	4	US-08-589-654-5	Sequence 5, App1
19	21.8	21.8	2022	5	PCT-US96-00996-4	Sequence 4, App1
20	21.6	21.6	1901	3	US-08-153-848-43	Sequence 43, App1
21	21.6	21.6	1225	3	US-08-494-577-11	Sequence 11, App1
22	21.6	21.6	6171	3	US-08-459-568-1	Sequence 1, App1
23	21.6	21.6	6171	4	US-08-399-411-1	Sequence 1, App1
24	21.6	21.6	1238	4	US-08-795-868-11	Sequence 11, App1
25	21.6	21.6	2793	4	US-08-795-868-13	Sequence 13, App1
26	21.6	21.6	1901	2	PCT-US93-11153-43	Sequence 43, App1
27	21.4	21.4	1816	5	US-08-061-889-1	Sequence 1, App1
28	21.4	21.4	1816	3	US-08-462-611-1	Sequence 1, App1
29	21.4	21.4	1915	3	US-08-496-855A-5	Sequence 5, App1
30	21.4	21.4	1224	4	US-08-210-762E-19	Sequence 19, App1
31	21.4	21.4	13351	4	US-08-210-762E-23	Sequence 23, App1
32	21.4	21.4	1915	4	US-08-466-589-11	Sequence 11, App1
33	21.4	21.4	10968	4	US-08-680-327-2	Sequence 2, App1
34	21.4	21.4	1816	5	PCT-US94-05378-1	Sequence 1, App1
35	21.2	21.2	7641	1	US-07-924-028A-5	Sequence 5, App1
36	21	21.0	13254	1	US-08-276-852-156	Sequence 156, App
37	21	21.0	13254	2	US-08-276-852-170	Sequence 170, App

ALIGNMENTS

c	38	21	21.0	13254	3	US-08-899-575-156	Sequence 156, App
c	39	21	21.0	13254	3	US-08-899-575-170	Sequence 170, App
c	40	21	21.0	13254	3	US-08-899-575-156	Sequence 156, App
c	41	21	21.0	13254	3	US-08-899-575-170	Sequence 170, App
c	42	21	21.0	9126	3	US-08-580-038-26	Sequence 26, App1
c	43	21	21.0	4080	4	US-08-710-249-3	Sequence 3, App1
c	44	21	21.0	1245	4	US-08-750-524-2	Sequence 2, App1
c	45	21	21.0	13254	5	PCT-US95-08743-170	Sequence 170, App

RESULT 1
US-08-466-589-7
Sequence 7, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Eliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & Mcclain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 73..1581
US-08-466-589-7
Query Match 83.4%; Score 83.4; DB 4; Length 1876;
Best Local Similarity 98.8%; Pred. No. 2,6e-22;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 9 AGTCGATATAGAGCGCTTGAAGCCACATTCACACATGATGATTCGAGG 68
|||||
Db 421 AGTCGATATAGAGCGCTTGAAGCCACATTCACACATGATGATTCGAGG 480
OY 69 CATTCGACATGACCTGCTCCAGGTA 93

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:17 : Search time 234.98 Seconds
(without alignments)
14.435 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37

Sequence: 1 ACCACACAGCATATTCAGAGTCTGCTACAT G 37

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

abase : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCITUS9.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	81.1	1876	4	Sequence 7, Appl1
2	21	56.8	2374	4	Sequence 5, Appl1
3	19.6	53.0	2277	3	Sequence 1, Appl1
4	19.6	53.0	2068	4	Sequence 1, Appl1
5	19.2	51.9	1299	4	Sequence 4, Appl1
6	18.8	50.8	812	5	Sequence 27, Appl1
7	18.2	49.2	1672	1	Sequence 28, Appl1
8	18.2	49.2	1845	1	Sequence 27, Appl1
9	18.2	49.2	1672	1	Sequence 28, Appl1
10	18.2	49.2	1845	1	Sequence 27, Appl1
11	18.2	49.2	1672	1	Sequence 28, Appl1
12	18.2	49.2	1845	1	Sequence 27, Appl1
13	17.8	48.1	2278	1	Sequence 6, Appl1
14	17.8	48.1	2415	1	Sequence 8, Appl1
15	17.8	48.1	2756	1	Sequence 12, Appl1
16	17.8	48.1	1980	2	Sequence 31, Appl1
17	17.8	48.1	7997	3	Sequence 14, Appl1
18	17.8	48.1	7997	3	Sequence 14, Appl1
19	17.8	48.1	264	4	Sequence 17, Appl1
20	17.8	48.1	350	5	Sequence 14, Appl1
21	17.8	48.1	264	5	Sequence 14, Appl1
22	17.4	47.0	5093	2	Sequence 14, Appl1
23	17.4	47.0	5093	2	Sequence 14, Appl1
24	17.4	47.0	3171	5	Sequence 14, Appl1
25	17.2	46.5	1133	1	Sequence 12, Appl1
26	17.2	46.5	1133	1	Sequence 12, Appl1
27	17.2	46.5	2400	1	Sequence 15, Appl1
28	17.2	46.5	2313	1	Sequence 15, Appl1
29	17.2	46.5	100	1	Sequence 16, Appl1
30	17.2	46.5	100	1	Sequence 16, Appl1
31	17.2	46.5	23	1	Sequence 25, Appl1
32	17.2	46.5	23	1	Sequence 25, Appl1
33	17.2	46.5	23	1	Sequence 4, Appl1
34	17.2	46.5	11531	1	Sequence 4, Appl1
35	17.2	46.5	23	1	Sequence 4, Appl1
36	17.2	46.5	11531	2	Sequence 4, Appl1
37	17.2	46.5	11561	3	Sequence 1, Appl1

38	17.2	46.5	1756	4	Sequence 3, Appl1
39	17.2	46.5	11561	4	Sequence 1, Appl1
40	17.2	46.5	23	4	Sequence 1, Appl1
41	17.2	46.5	305	5	Sequence 4, Appl1
42	17.2	45.9	33	3	Sequence 13, Appl1
43	17.2	45.9	531	4	Sequence 5, Appl1
44	17.2	45.9	531	4	Sequence 5, Appl1
45	16.8	45.4	17415	5	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-466-589-7
Sequence 7, Application US/08466589

Patent No. 5837489

GENERAL INFORMATION:

APPLICANT: Eliott, Kathryn J.

APPLICANT: Harpold, Michael M.

TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESS: Brown, Martin, Haller & McClain

STREET: 1660 Union Street

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92101-2926

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTA Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,589

FILING DATE: June 5, 1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/028,031

FILING DATE: March 8, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-9950

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

TELEX:

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1876 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 73..1861

US-08-466-589-7

Query Match

Best Local Similarity 100.0%; Pred. No. 0.00055;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 CAGGATATTCAGAGTCTGCTACATCG 37

Db 500 CAGGATATTCAGAGTCTGCTACATCG 529

RESULT 2
US-08-466-589-5
Sequence 5, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2374 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 184..2067
08-466-589-5

Query Match 56.8%; Score 21; DB 4; Length 2374;
Best Local Similarity 73.0%; Pred. No. 2.2;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCCAACAGCGCATATCAAGAGTTCCTGCTACATCG 37
DB 637 ACTCCCGCGCATTTACAGAGCTCTGCGAGCATCG 673

US-08-496-855A-1
Sequence 1, Application US/08496855A
Patent No. 5801232
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain

STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: U.S.A.
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,855A
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,503
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: 08-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9369B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 166..1755
US-08-496-855A-1

Query Match 53.0%; Score 19.6; DB 3; Length 2277;
Best Local Similarity 73.5%; Pred. No. 8.2;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CACACAGCGCATATTCAGAGTTCCTGCTACATCG 37
DB 688 CCCCCGCCATCTACAGAGCTCTGCGAGCATCG 721

US-08-466-589-1
Sequence 1, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:18; See ch time 234.98 Seconds
(without alignments)
152.93 Million cell updates/sec

Title: US-08-956-518a-101
Perfect score: 392
Sequence: 1 AGACGCAAGGAGGAGTAG..... TGACGCTCGGAGCTCAAC 392

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

base : Issued Patents NA:

- 1: /cgn2_6/prodata/1/lna/3A_COMB.seq:*
- 2: /cgn2_6/prodata/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/prodata/1/lna/5C_COMB.seq:*
- 4: /cgn2_6/prodata/1/lna/5D_COMB.seq:*
- 5: /cgn2_6/prodata/1/lna/PCITUS9_COMB.seq:*
- 6: /cgn2_6/prodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	72	18.4	1876	4	US-08-466-589-7	Sequence 7, Appl
2	56.4	14.4	12001	3	US-08-458-568A-11	Sequence 11, Appl
3	52.4	13.4	7218	2	US-08-232-563-14	Sequence 14, Appl
4	52.4	13.4	702	2	US-08-458-568A-3	Sequence 3, Appl
5	52.4	13.4	12001	3	US-08-458-568A-11	Sequence 11, Appl
6	51.8	13.2	4403	4	US-08-284-941-1	Sequence 1, Appl
7	51.8	13.2	4257	4	US-08-690-473-1	Sequence 1, Appl
8	51.8	13.2	4403	5	PCT-US93-02147A-1	Sequence 1, Appl
9	50.2	12.8	1335	5	PCT-US91-06532-1	Sequence 1, Appl
10	49	12.5	4405	1	US-07-885-972A-3	Sequence 3, Appl
11	49	12.5	223	1	US-08-383-761-1	Sequence 1, Appl
12	49	12.5	223	2	US-08-324-277-1	Sequence 1, Appl
13	48.4	12.3	7011	3	US-08-306-691B-42	Sequence 16, Appl
14	48.4	12.3	801	4	US-08-770-379-16	Sequence 4, Appl
15	47.8	12.2	1032	2	US-08-400-422-2	Sequence 2, Appl
16	47.4	12.1	1610	1	US-08-056-051-5	Sequence 5, Appl
17	47.4	12.1	803	1	US-07-928-611-12	Sequence 12, Appl
18	47.4	12.1	1610	1	US-07-928-611-21	Sequence 21, Appl
19	47.4	12.1	4425	4	US-08-749-169A-1	Sequence 1, Appl
20	47.4	12.1	803	4	US-08-487-811A-12	Sequence 12, Appl
21	47.4	12.1	1610	4	US-08-487-811A-21	Sequence 21, Appl
22	47.4	12.1	803	5	PCT-US93-07370-12	Sequence 12, Appl
23	47.4	12.1	1610	5	PCT-US93-07370-21	Sequence 21, Appl
24	46	11.7	18994	2	US-08-459-586-4	Sequence 4, Appl
25	46	11.7	18994	4	US-08-282-696-4	Sequence 4, Appl
26	45.8	11.7	11219	3	US-07-642-734C-1	Sequence 1, Appl
27	45.6	11.6	2588	4	US-08-796-414B-6	Sequence 6, Appl
28	44.8	11.4	2362	1	US-08-146-421-4	Sequence 4, Appl
29	44.6	11.4	30001	1	US-08-125-468-1	Sequence 1, Appl
30	44.6	11.4	2823	2	US-08-398-008A-1	Sequence 1, Appl
31	44.6	11.4	1288	3	US-08-440-856A-9	Sequence 9, Appl
32	44.6	11.4	30001	4	US-08-474-833-1	Sequence 1, Appl
33	44.4	11.3	6453	3	US-08-306-691B-14	Sequence 14, Appl
34	44.2	11.3	1723	1	US-07-841-646-28	Sequence 28, Appl
35	44.2	11.3	1723	1	US-07-901-703-10	Sequence 10, Appl
36	44.2	11.3	1723	1	US-08-147-023-28	Sequence 28, Appl
37	44.2	11.3	1723	1	US-08-206-864-3	Sequence 3, Appl

ALIGNMENTS

C 38	44.2	11.3	1723	2	US-08-278-729A-20	Sequence 20, Appl
C 39	44.2	11.3	1723	2	US-08-480-528A-7	Sequence 7, Appl
C 40	44.2	11.3	1723	2	US-08-479-666-7	Sequence 7, Appl
C 41	44.2	11.3	1723	2	US-08-155-343A-20	Sequence 20, Appl
C 42	44.2	11.3	1723	2	US-08-406-672-20	Sequence 20, Appl
C 43	44.2	11.3	1723	2	US-08-643-563A-20	Sequence 20, Appl
C 44	44.2	11.3	1723	2	US-08-447-570-28	Sequence 28, Appl
C 45	44.2	11.3	1723	5	PCT-US93-10520-7	Sequence 7, Appl

RESULT 1
US-08-466-589-7
Sequence 7, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & Mcclain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 336
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 73..1581
US-08-466-589-7
Query Match 18.4%; Score 72; DB 4; Length 1876;
Best Local Similarity 100.0%; Pred. No. 5e-06;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 GGCCGAGGCGGAGGCGGCGGAGGCGGAGGCGGCGGCTCGCTGAGCT 380
DB 1 GGCCGAGGCGGAGGCGGCGGCGGAGGCGGAGGCGGCGGCTCGCTGAGCT 60
QY 381 CCGGAGCTCAAC 392

Db 61 CCGGACTCAAC 72

```

1 RESULT 2
2 US-08-458-568A-11/C
3 Sequence 11, Application US/08458568A
4 Patent No. 5621339
5
6 GENERAL INFORMATION:
7 APPLICANT: Schaffer, Priscilla A.
8 APPLICANT: Yeh, Lily
9 TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
10 TITLE OF INVENTION: Infections
11 NUMBER OF SEQUENCES: 15
12
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
15 STREET: One Liberty Place, 46th floor
16 CITY: Philadelphia
17 STATE: PA
18 COUNTRY: USA
19 ZIP: 19103
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: Wordperfect 5.1
26
27 CURRENT APPLICATION DATA:
28 APPLICATION NUMBER: US/08/458,568A
29 FILING DATE: 02-JUNE-1995
30 CLASSIFICATION: 435
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/065,146
34 FILING DATE: 05-MAY-1993
35 CLASSIFICATION: 435
36
37 ATTORNEY/AGENT INFORMATION:
38 NAME: Leary Ph.D., Kathryn R.
39 REGISTRATION NUMBER: 36,317
40 REFERENCE/DOCKET NUMBER: DPCI-0029
41
42 TELECOMMUNICATION INFORMATION:
43 TELEPHONE: (215) 568-3100
44 TELEFAX: (215) 568-3439
45
46 INFORMATION FOR SEQ ID NO: 11:
47 SEQUENCE CHARACTERISTICS:
48 LENGTH: 12001 base pairs
49 TYPE: nucleic acid
50 STRANDEDNESS: double
51 TOPOLOGY: linear
52
53 MOLECULE TYPE: DNA (genomic)
54 HYPOTHETICAL: NO
55 ANTI-SENSE: NO
56 ORIGINAL SOURCE:
57 ORGANISM: Herpes simplex virus
58 STRAIN: Herpes Simplex Virus Type 1
59 US-08-458-568A-11

```

	Query Match	14.4%	Score 56.4	DB 3	Length 12001	
	Best Local Similarity	50.3%	Pred. No.	0.002		
	Matches 164	Conservative	0	Mismatches 111	Indels 1	Gaps 1
QY	43	GGCCTGCGCAGAGGGCCGAGGCCGAGAGCCCCTTGCTGTGA3ACTGGGGGATGCACTGCC	102			
Db	2319	GGGAGGGGGCGAAGGGCGGGAGGGGGCGGAGGGCGCGAGAGGGCGGAGGGGGG	2260			
QY	103	CGGAGGCTAACCCAGCCCCGGGAGTACCTCCGCTACACCTTGGGCTTCAGATTCCCTGGG	152			
Db	2259	CGAAGGGGCGGGAAGGGGGCGAGGGGGGGGAGGGGGGAGAGGGGCGGAGAGGGCG	2200			
QY	163	TGGCCGCGCAGACGCTGTGCCCGGGCTGGAGGATATCGCCGCGGGGAGCGGGGGGGGGG	222			
Db	2199	GGAAGGGGGCGAAGGGGGCGGAGGGGGCGAGAGGGCGGAGAGGGCGGAGAGGGGGG	2140			
QY	223	GGGGCTTGCTACGTGGAGAGGCGCGCGGGGGGCGGGGCGGGGGCGCGCGCCCGGCTT	282			

QY	Db	QY	Db
283	CTTAAAGCCCGCGAGACCCGAGCGGCGAGGTGCTCTGTGTGCCGCAAGCGCGAAGCCCCGGC	343	GACAGCCGACAGCTGAGAGCCCGCGG 368
2079	GAGGGGCGAGGGGGCGGAGGGGGCGACAGGGGGCGGAGGGGGCGG-AGGGGCGGAGGGGGCG	2020	GAGGGGCGGAGGGGGCGAGGGGGCGG 1995

```

1      RESULT      3
2      US-08-232-463-14/c
3      : Sequence 14, Application US/08232463
4      : Patent No. 5670367
5      : GENERAL INFORMATION:
6      : APPLICANT: DORNER, F.
7      : APPLICANT: SCHEIFLINGER, F.
8      : APPLICANT: FALKNER, F. G.
9      : TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
10     : NUMBER OF SEQUENCES: 52
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Foley & Lardner
13     : STREET: 1800 Diagonal Road, Suite 500
14     : CITY: Alexandria
15     : STATE: VA
16     : COUNTRY: USA
17     : ZIP: 22313-0299
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: PatentIn Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/232.463
25     : FILING DATE:
26     : CLASSIFICATION: 435
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US/07/935.313
29     : FILING DATE:
30     : APPLICATION NUMBER: EP 91 114 300.6
31     : FILING DATE: 26-AUG-1991
32     : ATTORNEY/AGENT INFORMATION:
33     : NAME: BENT, Stephen A.
34     : REGISTRATION NUMBER: 29,768
35     : REFERENCE/DOCKET NUMBER: 30472/114 IMMU
36     : TELECOMMUNICATION INFORMATION:
37     : TELEPHONE: (703)836-9300
38     : TELEFAX: (703)683-4109
39     : TELEX: 899149
40     : INFORMATION FOR SEQ. ID NO.: 14:
41     : SEQUENCE CHARACTERISTICS:
42     : LENGTH: 7218 base pairs
43     : TYPE: nucleic acid
44     : STRANDEDNESS: single
45     : TOPOLOGY: linear
46     : IMMEDIATE SOURCE:
47     : CLONE: PTZ9PC-F15
48     :
49     :
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :
101    :
102    :
103    :
104    :
105    :
106    :
107    :
108    :
109    :
110    :
111    :
112    :
113    :
114    :
115    :
116    :
117    :
118    :
119    :
120    :
121    :
122    :
123    :
124    :
125    :
126    :
127    :
128    :
129    :
130    :
131    :
132    :
133    :
134    :
135    :
136    :
137    :
138    :
139    :
140    :
141    :
142    :
143    :
144    :
145    :
146    :
147    :
148    :
149    :
150    :
151    :
152    :
153    :
154    :
155    :
156    :
157    :
158    :
159    :
160    :
161    :
162    :
163    :
164    :
165    :
166    :
167    :
168    :
169    :
170    :
171    :
172    :
173    :
174    :
175    :
176    :
177    :
178    :
179    :
180    :
181    :
182    :
183    :
184    :
185    :
186    :
187    :
188    :
189    :
190    :
191    :
192    :
193    :
194    :
195    :
196    :
197    :
198    :
199    :
200    :
201    :
202    :
203    :
204    :
205    :
206    :
207    :
208    :
209    :
210    :
211    :
212    :
213    :
214    :
215    :
216    :
217    :
218    :
219    :
220    :
221    :
222    :
223    :
224    :
225    :
226    :
227    :
228    :
229    :
230    :
231    :
232    :
233    :
234    :
235    :
236    :
237    :
238    :
239    :
240    :
241    :
242    :
243    :
244    :
245    :
246    :
247    :
248    :
249    :
250    :
251    :
252    :
253    :
254    :
255    :
256    :
257    :
258    :
259    :
260    :
261    :
262    :
263    :
264    :
265    :
266    :
267    :
268    :
269    :
270    :
271    :
272    :
273    :
274    :
275    :
276    :
277    :
278    :
279    :
280    :
281    :
282    :
283    :
284    :
285    :
286    :
287    :
288    :
289    :
290    :
291    :
292    :
293    :
294    :
295    :
296    :
297    :
298    :
299    :
300    :
301    :
302    :
303    :
304    :
305    :
306    :
307    :
308    :
309    :
310    :
311    :
312    :
313    :
314    :
315    :
316    :
317    :
318    :
319    :
320    :
321    :
322    :
323    :
324    :
325    :
326    :
327    :
328    :
329    :
330    :
331    :
332    :
333    :
334    :
335    :
336    :
337    :
338    :
339    :
340    :
341    :
342    :
343    :
344    :
345    :
346    :
347    :
348    :
349    :
350    :
351    :
352    :
353    :
354    :
355    :
356    :
357    :
358    :
359    :
360    :
361    :
362    :
363    :
364    :
365    :
366    :
367    :
368    :
369    :
370    :
371    :
372    :
373    :
374    :
375    :
376    :
377    :
378    :
379    :
380    :
381    :
382    :
383    :
384    :
385    :
386    :
387    :
388    :
389    :
390    :
391    :
392    :
393    :
394    :
395    :
396    :
397    :
398    :
399    :
400    :
401    :
402    :
403    :
404    :
405    :
406    :
407    :
408    :
409    :
410    :
411    :
412    :
413    :
414    :
415    :
416    :
417    :
418    :
419    :
420    :
421    :
422    :
423    :
424    :
425    :
426    :
427    :
428    :
429    :
430    :
431    :
432    :
433    :
434    :
435    :
436    :
437    :
438    :
439    :
440    :
441    :
442    :
443    :
444    :
445    :
446    :
447    :
448    :
449    :
450    :
451    :
452    :
453    :
454    :
455    :
456    :
457    :
458    :
459    :
460    :
461    :
462    :
463    :
464    :
465    :
466    :
467    :
468    :
469    :
470    :
471    :
472    :
473    :
474    :
475    :
476    :
477    :
478    :
479    :
480    :
481    :
482    :
483    :
484    :
485    :
486    :
487    :
488    :
489    :
490    :
491    :
492    :
493    :
494    :
495    :
496    :
497    :
498    :
499    :
500    :
501    :
502    :
503    :
504    :
505    :
506    :
507    :
508    :
509    :
510    :
511    :
512    :
513    :
514    :
515    :
516    :
517    :
518    :
519    :
520    :
521    :
522    :
523    :
524    :
525    :
526    :
527    :
528    :
529    :
530    :
531    :
532    :
533    :
534    :
535    :
536    :
537    :
538    :
539    :
540    :
541    :
542    :
543    :
544    :
545    :
546    :
547    :
548    :
549    :
550    :
551    :
552    :
553    :
554    :
555    :
556    :
557    :
558    :
559    :
560    :
561    :
562    :
563    :
564    :
565    :
566    :
567    :
568    :
569    :
570    :
571    :
572    :
573    :
574    :
575    :
576    :
577    :
578    :
579    :
580    :
581    :
582    :
583    :
584    :
585    :
586    :
587    :
588    :
589    :
590    :
591    :
592    :
593    :
594    :
595    :
596    :
597    :
598    :
599    :
600    :
601    :
602    :
603    :
604    :
605    :
606    :
607    :
608    :
609    :
610    :
611    :
612    :
613    :
614    :
615    :
616    :
617    :
618    :
619    :
620    :
621    :
622    :
623    :
624    :
625    :
626    :
627    :

```

	Query Match	13.4%	Score 52.4	DB 2	Length 7218	
	Best Local Similarity	3.6%	Pred. No.	0.014	Mismatches 133	Gaps 0;
	Matches	14	Conservative	21	Mismatches	133; Indels 0; Gaps 0;
OY	1	AAGACGCAAGGAGAAGTAGACCTGGCCTTGGCGACCCCTGGCCTGCCAGAGGCGG	60			
Db	1426	KRRR	1367			
OY	61	AGGCCGAGAGCCCGCTGGTGAGACTGGGGGTGAGAGTGC CGGAGCGTA CCGCGCC	120			
Db	1366	RRR	1307			

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:28 ; Search time 234.98 Seconds
(without alignments)
268.811 Million cell updates/sec

Title: US-08-956-518A-102
Perfect score: 689
Sequence: 1 AGCCCTTCCAGCGGTAG.....CAGAGTCTCTGCTACATCG 689

Scoring table: IDENTITY_NUC

Archived: 176461 seqs, 45838279 residues

base :
Issued Patents: NA.*
1: /cgn2_5/ptodata1/1na/5A_COMB.seq.*
2: /cgn2_5/ptodata1/1na/5B_COMB.seq.*
3: /cgn2_5/ptodata1/1na/5C_COMB.seq.*
4: /cgn2_5/ptodata1/1na/5D_COMB.seq.*
5: /cgn2_5/ptodata1/1na/PCTUS9_COMB.seq.*
6: /cgn2_5/ptodata1/1na/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	15.7	1876	4	US-08-466-589-7
2	40.2	5.8	2990	3	US-08-572-951-1
3	38	5.5	1578	3	US-08-681-129-1
4	36.6	5.3	3223	1	US-07-980-528-1
5	36.6	5.3	1915	3	US-08-496-855A-5
6	35.6	5.3	1915	4	US-08-466-589-11
7	35.4	5.1	1942	4	US-08-604-989A-11
8	35.2	5.1	3151	2	US-08-423-443B-1
9	34.8	5.1	12001	3	US-08-458-568A-11
10	34.6	5.0	2374	4	US-08-466-589-5
11	33.8	4.9	1018	1	US-08-444-083-6
12	33.8	4.9	1018	1	US-08-286-304-6
13	33.8	4.9	1018	1	US-08-442-745-6
14	33.8	4.9	1018	1	US-08-443-129-6
15	33.8	4.9	1018	2	US-08-443-952-6
16	33.8	4.9	1018	2	US-08-443-130-6
17	33.8	4.9	43280	2	US-08-804-227C-1
18	33.8	4.9	1018	5	PCT-US95-04467-6
19	33.6	4.9	8438	1	US-07-945-283-1
20	33.6	4.9	3068	1	US-07-984-044A-3
21	33.6	4.9	1597	1	US-08-166-316-1
22	33.6	4.9	3068	2	US-08-458-393-3
23	33.4	4.8	2064	2	US-08-343-428-1
24	33.4	4.8	1192	5	PCT-US93-06251-51
25	33.2	4.8	388	1	US-07-626-618A-1
26	33.2	4.8	388	1	US-08-333-977-1
27	33.2	4.8	2638	3	US-08-306-691B-46
28	33.2	4.8	4257	4	US-08-690-473-1
29	33.2	4.8	11219	4	US-07-642-734C-1
30	33.3	4.8	11558	5	PCT-US93-06251-23
31	32.8	4.8	2277	5	US-08-496-855A-2
32	32.6	4.7	1161	1	US-08-496-855A-2
33	32.6	4.7	1161	1	US-08-434-877-2
34	32.6	4.7	12001	3	US-08-458-568A-11
35	32.6	4.7	4257	4	US-08-690-473-1
36	32.4	4.7	2923	2	US-08-377-292-6
37	32.4	4.7	2923	4	US-07-989-847-7

ALIGNMENTS

C 38	32.2	4.7	11613	1	US-08-484-044-10	Sequence 10, Appl
C 39	32.2	4.7	1864	1	US-08-353-948-1	Sequence 1, Appl
C 40	32	4.6	6961	2	US-08-727-034-2	Sequence 2, Appl
C 41	32	4.6	3836	4	US-08-216-260-1	Sequence 1, Appl
C 42	31.8	4.6	2505	1	US-07-977-434-9	Sequence 9, Appl
C 43	31.8	4.6	2504	1	US-08-073-384C-3	Sequence 3, Appl
C 44	31.8	4.6	3070	1	US-08-428-732-3	Sequence 3, Appl
C 45	31.8	4.6	2504	1	US-08-254-359A-3	Sequence 3, Appl

RESULT 1
US-08-466-589-7
; Sequence 7, Application US/08466589
; Patent No. 5837489
; GENERAL INFORMATION:
; APPLICANT: Eliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,589
; FILING DATE: June 5, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: March 8, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9950
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..1581
; US-08-466-589-7
Query Match 15.7%; Score 108; DB 4; Length 1876;
Best Local Similarity 95.7%; Pred. No. 2.5e-21;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 574 CCGGATATGCTGATGAGCGGTTGAGCCACATTCACACTACGCTGCTAATTC 633
DB 414 CTAATACAGTCTGATGAGCGGTTGAGCCACATTCACACTACGCTGCTAATTC 473
QY 634 TTCTGGGATTCGCCAGTACTGCTCCAGGATATTCAGAGTCTCTCTACATCG 689

Db 474 TTCTGGGCGATTGCGAGTACTGCTGCTCCAGGCGATTCAGAGATTCTGCTACTGCG 529

RESULT 2
US-08-572-951-1
Sequence 1, Application US/08572951
Patent No. 5824790

GENERAL INFORMATION:
APPLICANT: KEELING, PETER L.
APPLICANT: KNIGHT, MARY E.
APPLICANT: GUAN, HANPING
TITLE OF INVENTION: MODIFICATION OF STARCH
TITLE OF INVENTION: SYNTHESIS IN PLANTS
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
ADDRESSEE: Intellectual Property Group of
STREET: Pillsbury Madison & Sutro LLP
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3918

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572,951
FILING DATE: 15-DEC-1995
CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/346,602
FILING DATE: 29-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/263,921
FILING DATE: 21-JUN-1994
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Paul N. Kokulis
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 222957/1.02.15C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2990 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-572-951-1

Query Match 5.8%; Score 40.2; DB 3; Length 2990;
Best Local Similarity 57.6%; Pred. N3: 0.025; Mismatches 53; Indels 0; Gaps 0;
Matches 72; Conservative 0;

QY 170 GCTCCACGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 229
Db 693 GCTCCACGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 752
QY 230 CGCGCGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 289
Db 753 CGCGCGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 812
QY 290 CGCGCG 294
Db 813 GCGCG 817

RESULT 3
US-08-681-129-1
Sequence 1, Application US/08681129
Patent No. 5738654

GENERAL INFORMATION:
APPLICANT: Mettenleiter, Thomas Cristoph
TITLE OF INVENTION: Pseudorabies virus vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330 Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,129
FILING DATE:
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/244,446
FILING DATE: 02-JUN-1994
CLASSIFICATION: 424
APPLICATION NUMBER: EP 92.203.079.6
FILING DATE: 06-OCT-1992

ATTORNEY/AGENT INFORMATION:
NAME: Gornley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1578 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
STRAIN: Kaplan
US-08-681-129-1

Query Match 5.5%; Score 38; DB 3; Length 1578;
Best Local Similarity 50.5%; Pred. N3: 0.08; Mismatches 90; Indels 0; Gaps 0;
Matches 92; Conservative 0;

QY 53 GCGGCTTGACGAGCGCGCGCTCTGCTGAGTGTATTAAGGAGAGAGCCCG 112
Db 1019 GCGGCTTGACGAGCGCGCGCTCTGAGTGTATTAAGGAGAGAGCCCG 1078
QY 113 CAGGCGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 172
Db 1079 GCGGCGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 1138
QY 173 CCGAGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 232
Db 1139 CCGAGAGGATGACGCGCGGAGAGTTGAGACCGCGAGAGCTCGCGCGGGGCGC 294
QY 233 CC 234
Db 1199 CC 1200

RESULT 4
US-07-980-528-1/c

SUMMARIES

XX	H50469;
AC	
XX	
SV	H50469.1
XX	

NI 6990310
 XX 22-SEP-1995 (Rel. 45, Created)
 DT 22-SEP-1995 (Rel. 45, Last updated, Version 1)
 XX
 DE Y029Gf08.r1 Homo sapiens cDNA 5' similar to contains MER22
 DE repetitive element ;
 XX
 XX EST.
 KW
 XX Homo sapiens (human)
 OS
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homidae; Homo.
 XX
 XX [1]
 RN 1-437
 RP
 RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,
 RA Hultman M., Kucaba T., Le M., Lennon G., Mearns J.,
 RA Rikkin L., Rohlfing T., Soares M., Tan F., Trevisan E., Waterston R.,
 RA Williamson A., Woldmann P., Wilson R.,
 RA "The Masur-Merck EST Project",
 RA Unpublished.
 CC
 CC Contact: Wilson R. Masur-Merck EST Project Washington University
 CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
 CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
 CC est@wustl.wustl.edu High quality sequence stops: 313 Source: IMAGE
 CC Consortium, LNU. This clone is available royalty-free through LNU
 CC ; contact the IMAGE Consortium (info@image.lnu.gov) for further
 CC Information. NCBI gi: 990310
 XX
 XX
 FH
 FH Key Location/Qualifiers
 FT source 1.437
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT mRNA <1..>437
 FT
 FT
 XX
 XX Sequence 437 BP; 67 A; 128 C; 127 G; 107 T; 8 other;

Query Match 12.7%; Score 7; DB 18; Length 437;
 Best Local Similarity 100.0%; Pred. No. 5.5e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 CTGCACG 55
 1111111
 328 CTGCACG 334

RESULT 2
 LOCUS D11490 480 bp mRNA EST 02-DEC-1992
 DEFINITION HM0000A22 Liver HepG2 cell line. Homo sapiens cDNA clone a-22.
 mRNA sequence.
 ACCESSION D11490
 NID 92148209
 VERSION D11490.1 GI:2148209
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 480)
 Okubo,K., Hori,N., Matoba,R., Miyama,T., Fukushima,A., Kojima,Y.
 and Matsubara,K.
 TITLE Large scale cDNA sequencing for analysis of quantitative and
 JOURNAL qualitative aspects of gene expression
 MEDLINE Nature Genet. 2, 173-179 (1992)
 COMMENT 94258199
 Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
 Miyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara

Institute for Molecular and Cellular Biology
 Osaka University
 1-3 Yamada-oka, Suita, Osaka 565, Japan.
 Location/Qualifiers
 1.480
 /organism="Homo sapiens"
 /db_xref="GDS:DO57502E"
 /db_xref="taxon:9606"
 /clone="a-22"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli."
 BASE COUNT 106 a 114 c 121 g 77 t 62 others
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 480;
 Best Local Similarity 100.0%; Pred. No. 5.5e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 CTGCACG 55
 1111111
 Db 137 CTGCACG 143

RESULT 3
 LOCUS D11514 383 bp mRNA EST 02-DEC-1992
 DEFINITION HM0000A20 Liver HepG2 cell line. Homo sapiens cDNA clone a20, mRNA
 sequence.
 ACCESSION D11514
 NID 92148192
 VERSION D11514.1 GI:2148192
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 383)
 Okubo,K., Hori,N., Matoba,R., Miyama,T., Fukushima,A., Kojima,Y.
 and Matsubara,K.
 TITLE Large scale cDNA sequencing for analysis of quantitative and
 JOURNAL qualitative aspects of gene expression
 MEDLINE Nature Genet. 2, 173-179 (1992)
 COMMENT 94258199
 Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
 Miyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
 Institute for Molecular and Cellular Biology
 1-3 Yamada-oka, Suita, Osaka 565, Japan.
 Location/Qualifiers
 1.383
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="a20"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli."
 BASE COUNT 106 a 106 c 104 g 59 t 8 others
 ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 CTGCACG 55
 1111111
 Db 51 CTGCACG 45

RESULT 4
 D11560 532 bp mRNA EST 02-DEC-1992
 LOCUS HUM0C12D08 Liver HepG2 cell line. Homo sapiens cDNA clone c12d08,
 DEFINITION mRNA sequence.
 ACCESSION D11560
 NID 92148706
 VERSION D11560.1 GI:2148706
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 532)
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
 and Matsubara,K.
 Large scale cDNA sequencing for analysis of quantitative and
 qualitative aspects of gene expression
 Nature Genet. 2, 173-179 (1992)
 94258199

REFERENCE
 AUTHORS

TITLE
 JOURNAL MEDLINE
 COMMENT

FEATURES
 source
 1..532
 /organism="Homo sapiens"
 /db_xref="GDB:D0580625"
 /db_xref="taxon:9606"
 /clone="c12d08"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli."
 and transformed into E.coli." 43 others

BASE COUNT 124 a 112 c 109 g 144 t

ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 532;
 Best Local Similarity 100.0%; Pred. No. 5.6e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
 |||||
 247 CTGCACG 253

RESULT 5
 D11631 376 bp mRNA EST 02-DEC-1992
 LOCUS HUM00C202 Liver HepG2 cell line. Homo sapiens cDNA clone c202,
 DEFINITION mRNA sequence.
 ACCESSION D11631
 NID 92148223
 VERSION D11631.1 GI:2148223
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 376)
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
 and Matsubara,K.
 Large scale cDNA sequencing for analysis of quantitative and
 qualitative aspects of gene expression
 Nature Genet. 2, 173-179 (1992)
 94258199

REFERENCE
 AUTHORS

TITLE
 JOURNAL MEDLINE
 COMMENT

FEATURES
 source
 1..411
 /organism="Homo sapiens"
 /db_xref="GDB:D058337E"
 /db_xref="taxon:9606"
 /clone="hm01902"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli."
 and transformed into E.coli." 2 others

BASE COUNT 116 a 119 c 109 g 65 t

ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 411;
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
 |||||
 Db 87 CTGCACG 81

CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
 Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
 Institute for Molecular and Cellular Biology
 Osaka University
 1-3 Yamada-oka, Suita, Osaka 565, Japan.

INSTITUTE FOR MOLECULAR AND CELLULAR BIOLOGY
 OSAKA UNIVERSITY
 1-3 YAMADA-OKA, SUIITA, OSAKA 565, JAPAN.
 LOCATION/QUALIFIERS
 1..376
 /organism="Homo sapiens"
 /db_xref="GDB:D057516E"
 /db_xref="taxon:9606"
 /clone="c202"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli." 57 t

BASE COUNT 107 a 107 c 105 g 57 t

ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 376;
 Best Local Similarity 100.0%; Pred. No. 5.3e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
 |||||
 Db 51 CTGCACG 45

RESULT 6
 D11791 411 bp mRNA EST 02-DEC-1992
 LOCUS HUM0M01G02 Liver HepG2 cell line. Homo sapiens cDNA clone hm01g02,
 DEFINITION mRNA sequence.
 ACCESSION D11791
 NID 92155069
 VERSION D11791.1 GI:2155069
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 411)
 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
 and Matsubara,K.
 Large scale cDNA sequencing for analysis of quantitative and
 qualitative aspects of gene expression
 Nature Genet. 2, 173-179 (1992)
 94258199

REFERENCE
 AUTHORS

TITLE
 JOURNAL MEDLINE
 COMMENT

FEATURES
 source
 1..411
 /organism="Homo sapiens"
 /db_xref="GDB:D058337E"
 /db_xref="taxon:9606"
 /clone="hm01902"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli." 2 others

BASE COUNT 116 a 119 c 109 g 65 t

ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 411;
 Best Local Similarity 100.0%; Pred. No. 5.4e+04;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
 |||||
 Db 87 CTGCACG 81

CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
 Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
 Institute for Molecular and Cellular Biology
 Osaka University
 1-3 Yamada-oka, Suita, Osaka 565, Japan.


```

RESULT 7
LOCUS D11889/c 404 bp mRNA EST 02-DEC-1992
DEFINITION HUMH05E09 Liver Hepg2 cell line. Homo sapiens CDNA clone hm05e09,
            mRNA sequence.
ACCESSION D11889
NID 92155149
VERSION D11889.1 GI:2155149
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 404)
Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
and Matsubara,K.
Large scale cDNA sequencing for analysis of quantitative and
qualitative aspects of gene expression
Nature Genet. 2, 173-179 (1992)
94258199

JRNAL
MEDLINE
COMMENT

FEATURES
Source
1. 404
/organism="Homo sapiens"
/db_xref="GDB:D058435E"
/db_xref="taxon:9606"
/clone="hm05e09"
/clone_1lb="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."

BASE COUNT 94 a 125 c 96 g 84 t 5 others

ORIGIN
Query Match 12.7%; Score 7; DB 20; Length 404;
Best Local Similarity 100.0%; Pred. No. 5.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
|||||
332 CTCGACG 326

RESULT 8
LOCUS D11893/c 244 bp mRNA EST 02-DEC-1992
DEFINITION HUMH05F08 Liver Hepg2 cell line. Homo sapiens CDNA clone hm05f08,
            mRNA sequence.
ACCESSION D11893
NID 92155151
VERSION D11893.1 GI:2155151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 244)
Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
and Matsubara,K.
Large scale cDNA sequencing for analysis of quantitative and
qualitative aspects of gene expression
Nature Genet. 2, 173-179 (1992)
94258199

JRNAL
MEDLINE
COMMENT

FEATURES
Source
1. 244
/organism="Homo sapiens"
/db_xref="GDB:D058435E"
/db_xref="taxon:9606"
/clone="hm05f08"
/clone_1lb="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."

BASE COUNT 134 a 105 c 109 g 101 t 9 others

ORIGIN
Query Match 12.7%; Score 7; DB 20; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.5e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
|||||
332 CTCGACG 326

Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki
Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
Institute for Molecular and Cellular Biology
Osaka University
1-3 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers

```

```

FEATURES
Source
1. 244
/organism="Homo sapiens"
/db_xref="GDB:D058435E"
/db_xref="taxon:9606"
/clone="hm05f08"
/clone_1lb="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."

BASE COUNT 69 a 72 c 49 g 51 t 3 others

ORIGIN
Query Match 12.7%; Score 7; DB 20; Length 244;
Best Local Similarity 100.0%; Pred. No. 5e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
|||||
171 CTCGACG 165

RESULT 9
LOCUS D12042/c 458 bp mRNA EST 02-DEC-1992
DEFINITION HUM0005155 Liver Hepg2 cell line. Homo sapiens CDNA clone s155,
            mRNA sequence.
ACCESSION D12042
NID 92148299
VERSION D12042.1 GI:2148299
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 458)
Okubo,K., Horii,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
and Matsubara,K.
Large scale cDNA sequencing for analysis of quantitative and
qualitative aspects of gene expression
Nature Genet. 2, 173-179 (1992)
94258199

JRNAL
MEDLINE
COMMENT

FEATURES
Source
1. 458
/organism="Homo sapiens"
/db_xref="GDB:D057600E"
/db_xref="taxon:9606"
/clone="s155"
/clone_1lb="Liver Hepg2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."

BASE COUNT 134 a 105 c 109 g 101 t 9 others

ORIGIN
Query Match 12.7%; Score 7; DB 20; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.5e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
|||||
171 CTCGACG 165

Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki
Niiyama, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara
Institute for Molecular and Cellular Biology
Osaka University
1-3 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers

```


Db 295 CTGCACG 289

RESULT 10

LOCUS

D12100/c

397 bp mRNA

EST

07-OCT-1996

HOMO0005173

Liver HepG2 cell line. Homo sapiens cDNA clone sl73 3',

DEFINITION

mRNA sequence.

ACCESSION

D12100

VERSION

9767711

D12100.1 GI:767711

KEYWORDS

EST.

SOURCE

human.

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

1 (bases 1 to 397)

AUTHORS

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y.

TITLE

Large scale cDNA sequencing for analysis of quantitative and

JOURNAL

Nature Genet. 2, 173-179 (1992)

MEDLINE

94258199

REFERENCE

2 (bases 1 to 397)

AUTHORS

Murakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.

TITLE

Chromosomal assignments of 3'-directed partial cDNA sequences

JOURNAL

representing novel genes expressed in granulocytoid cells

MEDLINE

95137584

COMMENT

Contact: Kouzaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki

FEATURES

Source

Location/Qualifiers

E COUNT

114 a

80 c

109 g

93 t

1 others

GIN

114 a

80 c

109 g

93 t

1 others

Query Match

12.7%; Score 7; DB 20; Length 397;

Best Local Similarity

100.0%; Pred. No. 5.4e+04;

Matches

7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y

49 CTGCACG 55

Db

187 CTGCACG 181

RESULT 11

D12355/c

356 bp mRNA

EST

02-DEC-1992

HOMO0007B04

Liver HepG2 cell line. Homo sapiens cDNA clone tb04,

DEFINITION

mRNA sequence.

ACCESSION

D12355

VERSION

62148499

D12355.1 GI:2148499

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

1 (bases 1 to 356)

AUTHORS

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y.

TITLE

Large scale cDNA sequencing for analysis of quantitative and

JOURNAL

Nature Genet. 2, 173-179 (1992)

MEDLINE

94258199

REFERENCE

2 (bases 1 to 356)

Contact: Kouzaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki

Institute for Molecular and Cellular Biology

Osaka University

1-3 Yamada-oka, Suita, Osaka 565, Japan.

Location/Qualifiers

1. 356

/organism="Homo sapiens"

/db_xref="GDB:D057837E"

/db_xref="taxon:9606"

/clone="td04"

/clone_lib="Liver HepG2 cell line."

/lab_host="E.coli"

/note="3'-directed regional cDNA library. Cleaved by MboI

and transformed into E.coli."

others

E COUNT

98 a

99 c

100 g

56 t

3 others

GIN

98 a

99 c

100 g

56 t

3 others

Query Match

12.7%; Score 7; DB 20; Length 356;

Best Local Similarity

100.0%; Pred. No. 5.3e+04;

Matches

7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y

49 CTGCACG 55

Db

51 CTGCACG 45

RESULT 12

D15140/c

322 bp mRNA

EST

20-JUL-1998

R1000143A

Rice callus Oryza sativa cDNA clone R, mRNA sequence.

DEFINITION

D15140

NID

9286332

VERSION

D15140.1 GI:286332

KEYWORDS

EST.

SOURCE

Oryza sativa.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Oryza.

REFERENCE

1 (bases 1 to 322)

AUTHORS

Sasaki, T. and Minobe, Y.

TITLE

Rice cDNA from callus

JOURNAL

Unpublished (1994)

COMMENT

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@agr.affrc.go.jp

PROJECT "RGP"

Seq primer: oligo(dT).

Location/Qualifiers

1. 322

/organism="Oryza sativa"

/strain="cultivar Nipponbare, sub-species Japonica"

/db_xref="taxon:4550"

/clone="R"

/clone_lib="Rice callus"

/note="Vector: pBluescript II SK+; Site 1: SalI; S'

NotI; cDNA prepared from rice callus mRNAs by us"

oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid. 10 others

BASE COUNT 53 a 83 c 85 g 91 t
ORIGIN
Query Match 12.7%; Score 7; DB 20; Length 322;
Best Local Similarity 100.0%; Pred. No. 5.2e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55
|||||||
Db 275 CTGCACG 269

RESULT 13
D15143 332 bp mRNA EST 20-JUL-1998
LOCUS D15143/c
DEFINITION RICC0149A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
SESSION D15143
9286335
D15143.1 GI:286335

KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Euphylliphytes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 332)
AUTHORS Sasaki, T. and Minobe, Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)

COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'

FEATURES
source
1..332
Location/Qualifiers

1..332
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone_id="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid.
BASE COUNT 69 a 85 c 96 g 80 t 2 others
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 332;
Best Local Similarity 100.0%; Pred. No. 5.3e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55
|||||||
Db 76 CTGCACG 70

RESULT 14
D15154 273 bp mRNA EST 20-JUL-1998
LOCUS D15154
DEFINITION RICC0177A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D15154

NID 9286345
VERSION D15154.1 GI:286345
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Euphylliphytes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 273)
AUTHORS Sasaki, T. and Minobe, Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)

COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'

FEATURES
source
1..273
Location/Qualifiers

1..273
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone_id="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid.
BASE COUNT 43 a 87 c 72 g 60 t 11 others
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 273;
Best Local Similarity 100.0%; Pred. No. 5.1e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 CTGCACG 55
|||||||
Db 85 CTGCACG 91

RESULT 15
D15173/c 356 bp mRNA EST 20-JUL-1998
LOCUS D15173
DEFINITION RICC0205A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D15173
NID 9286365
VERSION D15173.1 GI:286365
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Euphylliphytes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 356)
AUTHORS Sasaki, T. and Minobe, Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)

COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441

Fax: 0298-38-7468
Email: tsasakieabr.affrc.go.jp

PROJECT = "/RGP/
Seq primer: oligo(dT).

FEATURES
Source Location/Qualifiers

1. .356
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone_id="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 71 a 88 c 111 g 75 t 11 others
ORIGIN

Query Match 12.7%; Score 7; DB 20; Length 356;

Best Local Similarity 100.0%; Pred. No. 5.3e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
|||||
DB 102 CTGCACG 96

Search completed: September 17, 1999, 21:27:52
Job time: 14256 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:56 ; Search time 2825.05 Seconds

(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-87

Perfect score: 10

Sequence: 1 TGTGTCTCAG 10

Scoring table:

IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database:

EST*

1: em_est1:*\n2: em_est2:*\n3: em_est3:*\n4: em_est4:*\n5: em_est5:*\n6: em_est6:*\n7: em_est7:*\n8: em_est8:*\n9: em_est9:*\n10: em_est10:*\n11: em_est11:*\n12: em_est12:*\n13: em_est13:*\n14: em_est14:*\n15: em_est15:*\n16: em_est16:*\n17: em_est17:*\n18: em_est18:*\n19: em_est19:*\n20: em_est20:*\n21: em_est21:*\n22: em_est22:*\n23: em_est23:*\n24: em_est24:*\n25: em_est25:*\n26: em_est26:*\n27: em_est27:*\n28: em_est28:*\n29: em_est29:*\n30: em_est30:*\n31: em_est31:*\n32: em_est32:*\n33: em_est33:*\n34: em_est34:*\n35: em_est35:*\n36: em_est36:*\n37: em_est37:*\n38: em_est38:*\n39: em_est39:*\n40: em_est40:*\n41: em_est41:*\n42: em_est42:*\n43: em_est43:*\n44: em_est44:*\n45: em_est45:*\n46: em_est46:*\n47: em_est47:*\n48: em_est48:*\n49: em_est49:*\n50: em_est50:*\n51: em_est51:*\n52: em_est52:*\n53: em_est53:*

54: em_est22:*\n55: em_est23:*\n56: em_est24:*\n57: em_est25:*\n58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100.0	429	5	AF120344	AF120344 Homo sapi	
2	100.0	389	20	D12193	D12193 HUM000317	
3	100.0	566	20	D20010	D20010 HUM000981	
4	100.0	88	20	D20990	D20990 HUM001972	
5	100.0	199	20	D31098	D31098 HUM012503 H	
6	100.0	374	20	D32407	D32407 CELK0136R	
7	100.0	406	20	D33452	D33452 CELK031BR	
8	100.0	344	20	F02697	F02697 HSC17B112 r	
9	100.0	265	20	F03251	F03251 HSC1RA102 r	
10	100.0	314	20	M78805	M78805 EST00953 H	
11	100.0	290	20	M78839	M78839 EST00987 H	
12	100.0	332	20	M78856	M78856 EST01004 H	
13	100.0	390	20	M79816	M79816 WES000353 N	
14	100.0	225	20	M85650	M85650 EST02167 F	
15	100.0	345	20	M86021	M86021 EST02546 F	
16	100.0	247	20	M86024	M86024 EST02549 F	
17	100.0	526	20	M88899	M88899 CEL10FA Chr	
18	100.0	312	20	T02241	T02241 WES002962 E	
19	100.0	315	20	T02245	T02245 WES002966 E	
20	100.0	408	20	T07350	T07350 EST05239 Fe	
21	100.0	370	20	T08327	T08327 EST06218 I	
22	100.0	429	20	T08749	T08749 EST06641 I	
23	100.0	388	20	T09101	T09101 EST06994 I	
24	100.0	393	20	T11545	T11545 A1418F Hear	
25	100.0	214	20	T14734	T14734 05C04f11-I	
26	100.0	248	20	T15405	T15405 IB1093 Infa	
27	100.0	543	20	T15545	T15545 IB1523 Infa	
28	100.0	310	20	T16583	T16583 NIB1514 NOR	
29	100.0	415	20	T17119	T17119 NIB352 Norm	
30	100.0	337	20	T20645	T20645 2653 Lambda	
31	100.0	518	20	T26572	T26572 AB324D9F It	
32	100.0	450	20	T27084	T27084 NIB261G03R	
33	100.0	342	20	T28031	T28031 EST24388 Hu	
34	100.0	346	20	T28461	T28461 EST45746 Hu	
35	100.0	321	20	T29601	T29601 EST86422 Hu	
36	100.0	326	20	T30869	T30869 EST24325 Hu	
37	100.0	336	20	T32366	T32366 EST47555 Hu	
38	100.0	336	20	T32367	T32367 EST47556 Hu	
39	100.0	323	20	T33499	T33499 EST58071 Hu	
40	100.0	438	20	T33648	T33648 EST58566 Hu	
41	100.0	360	20	T34126	T34126 EST62996 Hu	
42	100.0	221	20	T34206	T34206 EST64188 Hu	
43	100.0	258	20	T35232	T35232 EST81801 Hu	
44	100.0	263	20	T35321	T35321 EST82821 Hu	
45	100.0	495	54	HSW009549	A1044699 Homo sapi	

ALIGNMENTS

RESULT 1
ID AF120344
XX AF120344 standard; RNA; EST; 429 BP.
AC AF120344;
XX
SV AF120344.1
XX

NI 94325197
 XX 11-MAR-1999 (Rel. 59, Created)
 DT 11-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX
 DE Homo sapiens clone IEDP1-29, mRNA sequence.
 XX
 KW EST.
 XX
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homnidae; Homo.
 XX
 XX [1]
 RP 1-429
 RA MEDLINE; 97471696.
 RA Jacob A.N., Baskaran N., Kandpal G., Narayan D., Bhargava A.K.,
 Kandpal R.P.;
 "Isolation of human ear specific cDNAs and construction of cDNA
 libraries from surgically removed small amounts of inner ear tissues";
 Somat. Cell Mol. Genet. 23(2):83-95(1997).
 XX
 XX [2]
 RP 1-429
 RA Nellisery J.K., Kandpal R.P.;
 Submitted (15-JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL Fels Institute for Cancer Research, Temple University School of
 Medicine, 3307 North Broad Street, Philadelphia, PA 19140, USA
 XX
 XX Key Location/Qualifiers
 FH 1.429
 FT source
 FT /db_xref="taxon:9606"
 FT /note="Isolated by kinetic enrichment in a representational
 difference analysis (RDA) experiment."
 FT /organism="Homo sapiens"
 FT /tissue-type="inner and middle ear"
 FT /dev_stage="fetus"
 FT /clone="IEDP1-29"
 FT
 FT
 XX
 SO Sequence 429 BP; 146 A; 58 C; 85 G; 134 T; 6 other;

Query Match 100.0%; Score 10; DB 5; Length 429;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TGTGTCTCAG 10
 244 TGTGTCTCAG 253

RESULT 2
 D12193
 LOCUS D12193 389 bp mRNA EST 07-OCT-1996
 DEFINITION HUM0005317 Liver HepG2 cell line. Homo sapiens CDNA clone s317 3',
 mRNA sequence.
 ACCESSION D12193
 MID 9767720
 VERSION D12193.1 GI:767720
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 389)
 AUTHORS Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.
 TITLE Large scale cDNA sequencing for analysis of quantitative and
 qualitative aspects of gene expression
 JOURNAL Nature Genet. 2, 173-179 (1992)
 MEDLINE 94258199

REFERENCE 2 (bases 1 to 389)
 AUTHORS Murakawa,K., Matsubara,K., Fukushima,A., Yoshii,J. and Okubo,K.
 TITLE Chromosomal assignments of 3'-directed partial cDNA sequences
 representing novel genes expressed in granulocytoid cells
 JOURNAL Genomics 23, 379-389 (1994)
 MEDLINE 95137584
 COMMENT Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
 Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara
 Institute for Molecular and Cellular Biology
 Osaka University
 1-3 Yamada-oka, Suita, Osaka 565, Japan.
 FEATURES
 source Location/Qualifiers
 1.389
 /organism="Homo sapiens"
 /db_xref="GDB:D087694E"
 /db_xref="taxon:9606"
 /map="14"
 /clone="s317"
 /clone_1lb="Liver HepG2 cell line."
 /lab_host="E.coli"
 /note="3'-directed regional cDNA library. Cleaved by MboI
 and transformed into E.coli."

BASE COUNT 101 a 94 c 70 g 124 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 389;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TGTGTCTCAG 10
 286 TGTGTCTCAG 295

RESULT 3
 D20010
 LOCUS D20010 566 bp mRNA EST 30-JUL-1996
 DEFINITION HUMG500981 Human promyelocyte Homo sapiens CDNA clone pm2779 3',
 mRNA sequence.
 ACCESSION D20010
 MID 9500907
 VERSION D20010.1 GI:500907
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 566)
 AUTHORS Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
 Yoshinari,H., Arimoto,J. and Matsubara,K.
 TITLE Gene expression of human promyelocytic cell line HL60 before and
 after induction of differentiation. A new application of 3'-directed
 cDNA sequencing
 JOURNAL Unpublished (1993)
 COMMENT Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
 Yoshinari,H., Arimoto,J. and Matsubara,K.
 Institute for Molecular and Cellular Biology
 Osaka University
 3-1 Yamada-oka, Suita, Osaka 565, Japan.
 FEATURES
 source Location/Qualifiers
 1.566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="pm2779"
 /clone_1lb="Human promyelocyte"
 /note="Female, adult, cell_line = HL60, cell_type =
 promyelocyte."

BASE COUNT 138 a 114 c 118 g 176 t 20 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 566;
 Best Local Similarity 100.0%; Pred. No. 5.6e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTGAG 10
 |||||
 Db 164 TGTGTGTGAG 173

RESULT 4
 D20990/c 88 bp mRNA EST 30-JUL-1996
 LOCUS HUMGS01972 Human promyelocyte Homo sapiens cDNA clone mp2519.3',
 DEFINITION mRNA sequence.
 ACCESSION D20990
 NID 9504810
 VERSION D20990.1 GI:504810
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 88)
 Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,
 Yoshinari, H., Arimoto, J. and Matsubara, K.,
 Gene expression of human promyelocytic cell line HL60 before and
 after induction of differentiation. A new application of 3'directed
 cDNA sequencing
 JOURNAL Unpublished (1993)
 COMMENT

FEATURES
 SOURCE
 1. 88
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="mp2519"
 /note="Female, adult, cell_line = HL60, cell_type =
 promyelocyte." 12 g 32 t 1 others

BASE COUNT 23 a 20 c 12 g 32 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTGAG 10
 |||||
 Db 52 TGTGTGTGAG 43

RESULT 5
 D31098 199 bp mRNA EST 08-FEB-1995
 LOCUS HUMH12503 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.
 DEFINITION
 ACCESSION D31098
 NID 9643978
 VERSION D31098.1 GI:643978
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 199)
 Sudo, K., Chinen, K. and Nakamura, Y.,
 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
 library

JOURNAL Genomics 24, 276-279 (1995)
 MEDLINE 95213017
 COMMENT
 Contact: Yusuke Nakamura
 Institute of Medical Science
 University of Tokyo
 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
 Tel: 81-3-5449-5372
 Fax: 81-3-5449-5433
 Email: yusuke@ims.u-tokyo.ac.jp
 Insert Length: 552 Std Error: 0.00
 High quality sequence stop: 1.
 Location/Qualifiers
 1. 199
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="17q21"
 /clone_lib="Human fetal lung"

BASE COUNT 58 a 40 c 36 g 59 t 6 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 199;
 Best Local Similarity 100.0%; Pred. No. 5.4e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTGAG 10
 |||||
 Db 162 TGTGTGTGAG 171

RESULT 6
 D32407 374 bp mRNA EST 08-AUG-1994
 LOCUS CELK01366R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone YK1366.3', mRNA sequence.
 ACCESSION D32407
 NID 9523368
 VERSION D32407.1 GI:523368
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
 Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 374)
 Kohara, Y., Mitsuki, H., Nishigaki, A., Motosashi, T., Sugimoto, A. and
 Tabara, H.,
 Toward an expression map of the C. elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

FEATURES
 SOURCE
 1. 374
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /map="978E04; 1"
 /clone="YK1366"
 /note="Yuj1 Kohara unpublished cDNA"
 /note="dev-stage=varied, sex=Hermaphrodite male,
 tissue=type=whole animal"
 Location/Qualifiers
 1. 374
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="mp2519"

BASE COUNT 119 a 70 c 67 g 118 t
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 374;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGTGTGTGAG 10
 Db 196 TGTGTGTGAG 205

RESULT 7

D33452 406 bp mRNA EST 08-AUG-1994
 LOCUS CEK031B8 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone YK31B8 3', mRNA sequence.

ACCESSION

D33452

NID

9524372

VERSION

D33452.1 GI:524372

KEYWORDS

EST.

SOURCE

Caenorhabditis elegans.

~ORGANISM

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiidae;
 Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; Caenorhabditis.
 1 (bases 1 to 406)

REFERENCE

Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
 Tabara,H.
 Toward an expression map of the C.elegans genome
 Unpublished (1994)

TITLE

On Sep 21, 1992 this sequence version replaced gi:279017.

JOURNAL

On Sep 21, 1992 this sequence version replaced gi:279017.

COMMENT

On Sep 21, 1992 this sequence version replaced gi:279017.

FEATURES

Location/Qualifiers

1..406

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db_xref="taxon:6239"

/clone="YK31B8"

/clone_lib="Yuj1 Kohara unpublished cDNA"

/note="dev-stage-varied, sex-Hermaphrodite male,
 tissue-type=whole animal"

tissue-type=whole animal"

BASE COUNT 162 a 40 c 100 g 104 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 406;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAG 10
 Db 328 TGTGTGTGAG 337

RESULT 8
 F02697/c 344 bp mRNA EST 02-FEB-1995
 LOCUS HSC17B112 normalized infant brain cDNA Homo sapiens cDNA clone
 DEFINITION c-17b11 3', mRNA sequence.
 ACCESSION F02697
 NID 9646254
 VERSION F02697.1 GI:646254
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 344)
 Aufiray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
 Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B.,
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,

TITLE Sebastiani-Kabakchis,C. and Tessier,A.
 JOURNAL IMAGE: molecular integration of the analysis of the human genome
 MEDLINE and its expression
 COMMENT C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534

Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read, removed at sequence 5'end
 Genexpress_library_id: C; Genexpress_sequence_id: alc-17b11
 Seq primer: (-21)M13-universal
 Location/Qualifiers
 1..344
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="10"
 /clone="c-17b11"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue-type="total brain"
 /dev-stage="3 months old"
 /note="Organ: brain; Vector: lambdafmd BA; Site_1: HindIII;
 Site_2: NotI; sex=Female; dev-stage=3 months old;
 isolate=muscular atrophy patient; tissue-type=total
 brain; total mRNA was oligo-(dt) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambdafmd BA vector. Clone library from B.Soures, Psychiatry
 Dept. Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S. in press"

BASE COUNT 115 a 71 c 55 g 101 t 2 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 344;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAG 10
 Db 45 TGTGTGTGAG 36

RESULT 9

F03251 265 bp mRNA EST 02-FEB-1995
 LOCUS HSC1A102 normalized infant brain cDNA Homo sapiens cDNA clone
 DEFINITION c-1ra10 3', mRNA sequence.
 ACCESSION F03251
 NID 9646808
 VERSION F03251.1 GI:646808
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 265)
 Aufiray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
 Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B.,
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
 Sebastiani-Kabakchis,C. and Tessier,A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534
 On Apr 14, 1993 this sequence version replaced.

Contact: Genethon
 Genexpress-Genethon

Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. removed at sequence 5' end
Genexpress library idt: C; Genexpress_sequence_idt: alc-lra10
Seq primer: (-21)M13-universal.

FEATURES

source

Location/Qualifiers

1. 265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q21"
/clone="c-lra10"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lambda BA; Site 1: HindIII;
Site 2: NotI; sex=Female; dev_stage=3 months old;
Isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lambda BA vector. Clone library from B. Soares, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S. in press"

BASE COUNT

78 a 33 c 71 g 75 t 8 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 265;
Best Local Similarity 100.0%; Pred. No. 5.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
|||||

Db 227 TGTGTGTGAC 236

RESULT 10

M78805 314 bp mRNA EST 26-MAY-1992
LOCUS EST00953 Hippocampus, Strata gene (cat. #936205) Homo sapiens cDNA
DEFINITION clone HHC602, mRNA sequence.

ACCESSION M78805
NID 9273120
VERSION M78805.1 GI:273120
KEYWORDS EST.
RCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 314)

TITLE Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 632-634 (1992)
MEDLINE 92168112
AUTHORS 2 (bases 1 to 314)
REFERENCE Polymeropoulos, M.H., Xiao, H., Sikela, J.M., Adams, M., Venter, J.C.
and Merrill, C.R.
Chromosomal distribution of 320 genes from a brain cDNA library
JOURNAL Nature Genet. 4, 381-386 (1993)
MEDLINE 94004966

COMMENT

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.

FEATURES

source

Location/Qualifiers

1. 314
/organism="Homo sapiens"
/db_xref="ATCC (inhost):78249"
/db_xref="GDB:D12S110E"
/db_xref="taxon:9606"
/map="12"
/clone="HHC602"

/clone_lib="Hippocampus, Strata gene (cat. #936205)"
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
average insert size."

BASE COUNT

70 a 67 c 80 g 94 t 3 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 314;
Best Local Similarity 100.0%; Pred. No. 5.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
|||||

Db 35 TGTGTGTGAC 26

RESULT 11

M78839 290 bp mRNA EST 26-MAY-1992
LOCUS EST00987 Hippocampus, Strata gene (cat. #936205) Homo sapiens cDNA
DEFINITION clone HHC695, mRNA sequence.

ACCESSION M78839
NID 9273154
VERSION M78839.1 GI:273154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 290)

TITLE Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 632-634 (1992)
MEDLINE 92168112
AUTHORS 2 (bases 1 to 290)
REFERENCE Polymeropoulos, M.H., Xiao, H., Sikela, J.M., Adams, M., Venter, J.C.
and Merrill, C.R.
Chromosomal distribution of 320 genes from a brain cDNA library
JOURNAL Nature Genet. 4, 381-386 (1993)
MEDLINE 94004966

COMMENT

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.

FEATURES

source

Location/Qualifiers

1. 290
/organism="Homo sapiens"
/db_xref="ATCC (inhost):78292"
/db_xref="GDB:D10S238E"
/db_xref="taxon:9606"
/map="10"

/clone="HHC695"
/clone_lib="Hippocampus, Strata gene (cat. #936205)"
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
average insert size."

BASE COUNT

76 a 99 c 52 g 61 t 2 others

Query Match 100.0%; Score 10; DB 20; Length 290;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10
 |||||
 Db 245 TGTGTCTCAG 236

RESULT 12
 M78856 332 bp mRNA EST 26-MAY-1992
 LOCUS M78856/c
 DEFINITION EST01004 Hippocampus, Striatogene (cat. #936205) Homo sapiens CDNA
 clone HRCMH72, mRNA sequence.
 ACCESSION M78856
 9273171
 ION M78856.1 GI:273171
 .WORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 332)
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 forward.
 Location/Qualifiers
 1. 332
 /organism="Homo sapiens"
 /db_xref="Arcc (Inhost):78321"
 /db_xref="GDB:DS1837E"
 /db_xref="taxon:9606"
 /clone="HRCMH72"
 /clone_lib="Hippocampus, Striatogene (cat. #936205)"
 /note="Vector: lambdaZAP-II; Female, 2 years; oligo-dt +
 random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
 average insert size."
 BASE COUNT 91 a 111 c 59 g 71 t
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 332;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10
 |||||
 Db 107 TGTGTCTCAG 98

RESULT 13
 M79816/c 390 bp mRNA EST 30-JUN-1992
 LOCUS M79816
 DEFINITION EST00353 Mixed stage, Striatogene (cat. #937006) Caenorhabditis
 elegans CDNA clone CEMSE23, mRNA sequence.
 ACCESSION M79816
 9271835
 NID M79816.1 GI:271835
 VERSION EST.
 KEYWORDS
 SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 390)
 AUTHORS McCombie,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,
 Uterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and
 Fields,C.
 TITLE Caenorhabditis elegans expressed sequence tags identify gene
 families and potential disease gene homologues
 JOURNAL Nature Genet. 1, 124-131 (1992)
 MEDLINE 93250983
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 forward.
 Location/Qualifiers
 1. 390
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="CEMSE23"
 /clone_lib="Mixed stage, Striatogene (cat. #937006)"
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage CDNA
 library. Striatogene catalog #937006. The library is oligo
 dt primed and directionally cloned in the Uni-ZAP XR
 vector."
 BASE COUNT 127 a 67 c 70 g 120 t 6 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 390;
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTCTCAG 10
 |||||
 Db 139 TGTGTCTCAG 130

RESULT 14
 M85650 225 bp mRNA EST 26-MAY-1992
 LOCUS M85650/c
 DEFINITION EST02167 Fetal brain, Striatogene (cat.#936206) Homo sapiens CDNA
 clone HBC138, mRNA sequence.
 ACCESSION M85650
 9274297
 NID M85650.1 GI:274297
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 225)
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 forward.
 Location/Qualifiers
 1. 225

Search completed: September 17, 1999, 21:27:57
 Job time: 14261 sec

```

/organism="Homo sapiens"
/db_xref="ATCC (lnhost):81227"
/db_xref="GDB:ID05343E"
/db_xref="taxon:9606"
/clone="HFBCL38"
/clone_1lb="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insert size."
BASE COUNT      67 a      67 c      49 g      41 t      1 others
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 225;
Best Local Similarity 100.0%; Pred. No. 5.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGTGTCTCAG 10
|||||
66 TGTGTCTCAG 57

```

```

RESULT 15
M86021      345 bp      mRNA      EST      26-MAY-1992
LOCUS      EST02546 Fetal brain, Stratagene (cat#936206) Homo sapiens CDNA
DEFINITION clone HFB0Y34, mRNA sequence.
ACCESSION  M86021
NID        9274672
VERSION    M86021.1 GI:274672
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 345)
AUTHORS   Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
            Sequence identification of 2,375 human brain genes
            Nature 355, 632-634 (1992)
            92168112
JOURNAL
MEDLINE
COMMENT    Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: akkerlav@igt.org
            Seq primer: M13 Forward.
            Location/Qualifiers

```

```

FEATURES
    source
        1..345
            /organism="Homo sapiens"
            /db_xref="ATCC (lnhost):81598"
            /db_xref="GDB:ID05709E"
            /db_xref="taxon:9606"
            /clone="HFB0Y34"
            /clone_1lb="Fetal brain, Stratagene (cat#936206)"
            /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
            oligo-dt + random primed cDNA synthesis; lambdaZAP-II
            vector, 1.0kb average insert size."
BASE COUNT      110 a      68 c      63 g      104 t
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 345;
Best Local Similarity 100.0%; Pred. No. 5.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGTGTCTCAG 10
|||||
DB      18 TGTGTCTCAG 27

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:55 : Search time 2825.05 Seconds
(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-86
Perfect score: 10
Sequence: 1 TTTTGAAG 10

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database:

EST:
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: gb_est1.*
21: gb_est2.*
22: gb_est3.*
23: gb_est4.*
24: gb_est5.*
25: gb_est6.*
26: gb_est7.*
27: gb_est8.*
28: gb_est9.*
29: gb_est10.*
30: gb_est11.*
31: gb_est12.*
32: gb_est13.*
33: gb_est14.*
34: gb_est15.*
35: gb_est16.*
36: gb_est17.*
37: gb_est18.*
38: gb_est19.*
39: gb_est20.*
40: gb_est21.*
41: gb_est22.*
42: gb_est23.*
43: gb_est24.*
44: gb_est25.*
45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: em_est20.*
53: em_est21.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	10	100.0	540	5 AF123245	AF123245 Gossypium
C 2	10	100.0	314	20 D20170	D20170 HUNG501144
C 3	10	100.0	288	20 D23028	D23028 RICC10206A
C 4	10	100.0	328	20 D22533	D22533 RICC01908 R
C 5	10	100.0	261	20 D23618	D23618 RICC3103A R
C 6	10	100.0	331	20 D24868	D24868 RICR2677A R
C 7	10	100.0	423	20 D26922	D26922 CELK014D2R
C 8	10	100.0	398	20 D27177	D27177 CELK004D5R
C 9	10	100.0	471	20 D27179	D27179 CELK002GXR
C 10	10	100.0	461	20 D27520	D27520 CELK012B4R
C 11	10	100.0	417	20 D27544	D27544 CELK013A7R
C 12	10	100.0	403	20 D27856	D27856 CELK005B2R
C 13	10	100.0	387	20 D27881	D27881 CELK005GXF
C 14	10	100.0	470	20 D27882	D27882 CELK005GXR
C 15	10	100.0	403	20 D27887	D27887 CELK005G2R
C 16	10	100.0	365	20 D28006	D28006 CELK007F6F
C 17	10	100.0	445	20 D28019	D28019 CELK008A1R
C 18	10	100.0	300	20 D31015	D31015 HUN112171 H
C 19	10	100.0	382	20 D32317	D32317 CELK010E5R
C 20	10	100.0	352	20 D32380	D32380 CELK012D2R
C 21	10	100.0	417	20 D32546	D32546 CELK018E8R
C 22	10	100.0	397	20 D32581	D32581 CELK019A6R
C 23	10	100.0	323	20 D32642	D32642 CELK019H3R
C 24	10	100.0	358	20 D32679	D32679 CELK001F4R
C 25	10	100.0	269	20 D32686	D32686 CELK020A5R
C 26	10	100.0	374	20 D32706	D32706 CELK020C9R
C 27	10	100.0	402	20 D32964	D32964 CELK024B7R
C 28	10	100.0	382	20 D32972	D32972 CELK024C5R
C 29	10	100.0	295	20 D33053	D33053 CELK025E2R
C 30	10	100.0	380	20 D33071	D33071 CELK025G7R
C 31	10	100.0	382	20 D33081	D33081 CELK026A9R
C 32	10	100.0	373	20 D33087	D33087 CELK026A9R
C 33	10	100.0	395	20 D33182	D33182 CELK027D9R
C 34	10	100.0	355	20 D33437	D33437 CELK031A9R
C 35	10	100.0	373	20 D33477	D33477 CELK031FXR
C 36	10	100.0	410	20 D33502	D33502 CELK031H7R
C 37	10	100.0	416	20 D33820	D33820 CELK036F3R
C 38	10	100.0	378	20 D33943	D33943 CELK038E9R
C 39	10	100.0	412	20 D33948	D33948 CELK038F4R
C 40	10	100.0	388	20 D34048	D34048 CELK040A8R
C 41	10	100.0	387	20 D34110	D34110 CELK041B4R
C 42	10	100.0	406	20 D34303	D34303 CELK044A3R
C 43	10	100.0	224	20 D34370	D34370 CELK045A2R
C 44	10	100.0	380	20 D34392	D34392 CELK045C6R
C 45	10	100.0	448	54 HSM009362	A1044512 Homo sapi

ALIGNMENTS

RESULT 1
ID AF123245/c standard; RNA; EST; 540 BP.
XX AF123245
NC AF123245:
XX
SV AF123245.1
XX

NI 94325284
 XX 11-MAR-1999 (Rel. 59, Created)
 DT 11-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX
 DE Gossypium hirsutum clone CFI6S strain Texas Marker-1, mRNA
 DE sequence.
 XX
 KW EST.
 XX
 OS Gossypium hirsutum (upland cotton)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 XX
 RN [1]
 RP 1-540
 PA Cui X., Brown R.M.;
 "CDNA sequences from secondary-stage cotton fibers";
 unpublished.
 XX
 RN [2]
 RP 1-540
 RA Cui X., Brown R.M.;
 ;
 RT Submitted (23-JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL Botany, University of Texas at Austin, Austin, TX 78713, USA
 XX
 FH Key Location/Qualifiers
 FT 1. 540
 FT /db_xref="taxon:3635"
 FT /organism="Gossypium hirsutum"
 FT /strain="Texas Marker-1"
 FT /tissue-type="24 days postanthesis secondary-stage fibers"
 FT /clone="CFI6S"
 FT /cell-type="seed epidermal cells"
 XX
 SQ Sequence 540 BP; 157 A; 105 C; 113 G; 165 T; 0 other;
 Query Match 100.0%; Score 10; DB 5; Length 540;
 Best Local Similarity 100.0%; Pred. No. 2.7e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OV 1 TTTTGGAG 10
 ||||||||
 525 TTTTGGAG 516
 RESULT 2
 D20170/c 314 bp mRNA EST 30-JUL-1996
 LOCUS D20170 Human promyelocyte Homo sapiens CDNA clone pm1519.3,
 DEFINITION mRNA sequence.
 ACCESSION D20170
 MID 9501267
 VERSION D20170.1 GI:501267
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 314)
 Murtakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.
 Chromosomal assignments of 3'-directed partial cDNA sequences
 representing novel genes expressed in granulocytoid cells
 Genomics 23, 379-389 (1994)
 JOURNAL 95137584
 MEDLINE 2 (bases 1 to 314)
 Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,
 Yoshinari, H., Arimoto, J. and Matsubara, K.
 Gene expression of human promyelocytic cell line HL60 before and
 TITLE

JOURNAL
 COMMENT after induction of differentiation. A new application of 3' directed
 cDNA sequencing
 unpublished (1993)
 CONTACT: Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,
 Yoshinari, H., Arimoto, J. and Matsubara, K.
 Institute for Molecular and Cellular Biology
 Osaka University
 3-1 Yamada-oka, Suita, Osaka 565, Japan.
 FEATURES
 SOURCE Location/Qualifiers
 1. 314
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="6"
 /clone="pm1519"
 /clone_lib="Human promyelocyte"
 /note="Female, adult, cell_line = HL60, cell_type =
 promyelocyte."
 BASE COUNT 91 a 59 c 66 g 83 t 15 others
 ORIGIN
 Query Match 100.0%; Score 10; DB 20; Length 314;
 Best Local Similarity 100.0%; Pred. No. 3.3e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OV 1 TTTTGGAG 10
 ||||||||
 DB 140 TTTTGGAG 131
 RESULT 3
 D22028/c 288 bp mRNA EST 20-JUL-1998
 LOCUS D22028 Rice callus Oryza sativa CDNA clone R, mRNA sequence.
 DEFINITION RICC10206A Rice callus Oryza sativa CDNA clone R, mRNA sequence.
 ACCESSION D22028
 MID 9426169
 VERSION D22028.1 GI:426169
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 288)
 Sasaki, T. and Minobe, Y.
 Rice cDNA from callus
 Unpublished (1994)
 COMMENT
 CONTACT: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0288-38-7441
 Fax: 0288-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT "RGP".
 FEATURES
 SOURCE Location/Qualifiers
 1. 288
 /organism="Oryza sativa"
 /strain="cultivar Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone="R"
 /clone_lib="Rice callus"
 /note="Vector: pBluescript II SK+; Site.1: SalI; Site.2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid."
 BASE COUNT 78 a 57 c 59 g 92 t 2 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 288;
 Best Local Similarity 100.0%; Pred. No. 3.4e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGGAG 10
 |||||
 DB 201 TTTTGGAG 192

RESULT 4
 D22533 328 bp mRNA EST 20-JUL-1998
 LOCUS D22533/c
 DEFINITION RICC0190B Rice callus Oryza sativa cDNA clone R, mRNA sequence.
 ACCESSION D22533
 NID 9425855
 VERSION D22533.1 GI:425855
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 328)
 /clone_1lb-"Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 PROJECT -"RGP".

REFERENCE
 AUTHORS Sasaki,T. and Minobe,Y.
 TITLE Rice cDNA from callus
 JOURNAL Unpublished (1994)

COMMENT
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp

BASE COUNT
 ORIGIN Location/Qualifiers
 1. 328
 /organism="Oryza sativa"
 /strain="cultivar Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone="R"
 /clone_1lb-"Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 PROJECT -"RGP".

FEATURES
 source

Query Match 100.0%; Score 10; DB 20; Length 328;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGGAG 10
 |||||
 DB 319 TTTTGGAG 310

RESULT 5
 D23618 261 bp mRNA EST 20-JUL-1998
 LOCUS D23618/c
 DEFINITION RICC3103A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
 ACCESSION D23618
 NID 9427550
 VERSION D23618.1 GI:427550
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 261)
 /clone_1lb-"Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 PROJECT -"RGP".

REFERENCE
 AUTHORS Sasaki,T. and Minobe,Y.
 TITLE Rice cDNA from callus
 JOURNAL Unpublished (1994)

COMMENT
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp

BASE COUNT
 ORIGIN Location/Qualifiers
 1. 261
 /organism="Oryza sativa"
 /strain="cultivar Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone="R"
 /clone_1lb-"Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 PROJECT -"RGP".

FEATURES
 source

Query Match 100.0%; Score 10; DB 20; Length 261;
 Best Local Similarity 100.0%; Pred. No. 3.5e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGGAG 10
 |||||
 DB 251 TTTTGGAG 242

RESULT 6
 D24868 331 bp mRNA EST 20-JUL-1998
 LOCUS D24868/c
 DEFINITION RICC2677A Rice root Oryza sativa cDNA clone R, mRNA sequence.
 ACCESSION D24868
 NID 9428776
 VERSION D24868.1 GI:428776
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 331)
 /clone_1lb-"Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 PROJECT -"RGP".

REFERENCE
 AUTHORS Minobe,Y. and Sasaki,T.
 TITLE Rice cDNA from root
 JOURNAL Unpublished (1995)

COMMENT
 Contact: Yuzo Minobe
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: minobe@rics0.riken.go.jp

BASE COUNT
 ORIGIN Location/Qualifiers
 1. 331
 /organism="Oryza sativa"
 /strain="Nipponbare, sub-species Japonica"

/db_xref="taxon:4530"
 /clone="R"
 /clone_lib="Rice root"
 /note="Prepared from seedling root."

BASE COUNT 89 a 63 c 68 g 109 t 2 others
 ORIGIN
 Query Match 100.0%; Score 10; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTTGAAG 10
 Db 166 TTTTGAAG 157

FEATURES 7
 922/c
 --CDS 423 bp mRNA EST 20-NOV-1995
 DEFINITION CELK01AD2R Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK14d2 3', mRNA sequence.
 ACCESSION D26922
 NID 9522008
 VERSION D26922.1 GI:522008
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 423)
 Kohara,Y., Mitsuki,H., Nishigaki,A., Mochashl,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1. 423
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /clone="YK14d2"
 /clone_lib="Yui Kohara unpublished cDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue=whole animal."
 BASE COUNT 160 a 52 c 104 g 107 t
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 423;
 Best Local Similarity 100.0%; Pred. No. 2.9e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTTGAAG 10
 Db 404 TTTTGAAG 395

RESULT 8
 D27177 398 bp mRNA EST 20-NOV-1995
 LOCUS
 DEFINITION CELK004D5R Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK4d5 3', mRNA sequence.
 D27177
 ACCESSION
 NID 9521441
 VERSION D27177.1 GI:521441

KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 398)
 Kohara,Y., Mitsuki,H., Nishigaki,A., Mochashl,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

FEATURES
 SOURCE
 Contact: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1. 398
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /clone="YK4d5"
 /clone_lib="Yui Kohara unpublished cDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue=whole animal."
 BASE COUNT 162 a 42 c 88 g 105 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 398;
 Best Local Similarity 100.0%; Pred. No. 3e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTTGAAG 10
 Db 230 TTTTGAAG 239

RESULT 9
 D27179 471 bp mRNA EST 20-NOV-1995
 LOCUS
 DEFINITION CELK002GXR Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK9g10 3', mRNA sequence.
 D27179
 ACCESSION
 NID 9521341
 VERSION D27179.1 GI:521341
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 471)
 Kohara,Y., Mitsuki,H., Nishigaki,A., Mochashl,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1. 471
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /clone="YK2g10"
 /clone_lib="Yui Kohara unpublished cDNA"

BASE COUNT 184 a 70 c 99 g 118 t
 ORIGIN /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

Query Match 100.0%; Score 10; DB 20; Length 471;
 Best Local Similarity 100.0%; Pred. No. 2.8e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10
 |||||||||
 Db 98 TTTTGAAG 107

RESULT 10
 D27520 461 bp mRNA EST 20-NOV-1995
 LOCUS CELK012B4R Yui Kohara unpublished CDNA Caenorhabditis elegans CDNA

DEFINITION clone yk12b4 3', mRNA sequence.
 D27520
 JSSION 9523006
 VERSION D27520.1 GI:523006
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.

ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 461)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1..461
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk12b4"
 /clone_id="Yui Kohara unpublished CDNA"
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

FEATURES
 source
 144 a 73 c 89 g 155 t

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 461;
 Best Local Similarity 100.0%; Pred. No. 2.8e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10
 |||||||||
 Db 226 TTTTGAAG 235

RESULT 11
 D27544 417 bp mRNA EST 20-NOV-1995
 LOCUS CELK013A7R Yui Kohara unpublished CDNA Caenorhabditis elegans CDNA
 DEFINITION clone yk13a7 3', mRNA sequence.
 D27544
 ACCESSION D27544
 NID 9523039
 VERSION D27544.1 GI:523039
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 417)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1..417
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk13a7"
 /clone_id="Yui Kohara unpublished CDNA"
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

BASE COUNT 112 a 85 c 75 g 144 t 1 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.9e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGAAG 10
 |||||||||
 Db 296 TTTTGAAG 287

RESULT 12
 D27856 403 bp mRNA EST 20-NOV-1995
 LOCUS CELK005B2R Yui Kohara unpublished CDNA Caenorhabditis elegans CDNA
 DEFINITION clone yk5b2 3', mRNA sequence.
 D27856
 ACCESSION D27856
 NID 9522587
 VERSION D27856.1 GI:522587
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.

ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 403)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1..403
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk5b2"
 /clone_id="Yui Kohara unpublished CDNA"
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

FEATURES
 source
 112 a 85 c 75 g 144 t 1 others

BASE COUNT 144 a 67 c 83 g 109 t
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 403;
Best Local Similarity 100.0%; Pred. No. 3e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGAAG 10
|||||
Db 127 TTTTGAAG 118

RESULT 13
D27881 387 bp mRNA EST 20-NOV-1995
LOCUS CELK005GAF Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk5g10 5', mRNA sequence.
TSSION D27881
9522636
D27881.1 GI:522636

VERSION D27881.1 GI:522636
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiidae; Rhabditiidae; Rhabditiidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 387)
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT

FEATURES
source Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..387
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk5g10"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 103 a 84 c 118 t 1 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 387;
Best Local Similarity 100.0%; Pred. No. 3e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGAAG 10
|||||
Db 270 TTTTGAAG 279

RESULT 14
D27882 470 bp mRNA EST 20-NOV-1995
LOCUS CELK005GCR Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk5g10 3', mRNA sequence.
ACCESSION D27882
9522637
D27882.1 GI:522637
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiidae; Rhabditiidae;

REFERENCE Rhabditiidae; Rhabditiidae; Rhabditiidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 470)
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT

Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..470
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk5g10"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

FEATURES
source

BASE COUNT 146 a 96 c 99 g 126 t 3 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 470;
Best Local Similarity 100.0%; Pred. No. 2.8e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTTGAAG 10
|||||
Db 191 TTTTGAAG 182

RESULT 15
D27887 403 bp mRNA EST 20-NOV-1995
LOCUS CELK005G2F Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk5g2 5', mRNA sequence.
ACCESSION D27887
9522624
D27887.1 GI:522624
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiidae; Rhabditiidae; Rhabditiidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 403)
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT

Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..403
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk5g2"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

FEATURES
source

BASE COUNT 154 a 63 c 76 g 110 t
ORIGIN

Sun Sep 19 10:26:32 1999

us-08-956-518a-86.rst

Page 7

Query Match 100.0%; Score 10; DB 20; Length 403;
Best Local Similarity 100.0%; Pred. No. 3e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TTTTGAAG 10
|||||||
Db 69 TTTTGAAG 60

Search completed: September 17, 1999, 21:27:56
Job time: 14260 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:26; Search time 425.19 Seconds

(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518a-85

Perfect score: 10

Sequence: 1 TCTCCTTAG 10

Scoring table: IDENTITY_NUC

Archived: 311585 seqs, 125096042 residues

--cbase: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	10	100.0	1 N90613	CD20 CDNA. Rapid 1
2	10	100.0	1 N90610	CD20.4 CDNA. Rapid
3	10	100.0	1 N90225	Malaria-specific P
4	10	100.0	1 N90732	Sequence of plasmid
5	10	100.0	1 N90733	Genomic DNA from P
6	10	100.0	1 N91650	DNA encoding Sn-gi
7	10	100.0	1 N92837	CDNA from Plasmid
8	10	100.0	1 N92837	Probe containing c
9	10	100.0	1 N81290	New recombinant CD
10	10	100.0	1 N71032	Encodes B. t. t inse
11	10	100.0	1 N71032	Sequence of synthe
12	10	100.0	1 N71032	Sequence contg. ab
13	10	100.0	1 N71032	Plasmidum virax 6
14	10	100.0	1 N71032	Clone 9C encoding
15	10	100.0	1 N71032	Urate oxidase-enco
16	10	100.0	1 N71032	Urate oxidase-enco
17	10	100.0	1 N71032	Promoter of nod1ab
18	10	100.0	1 N71032	Encodes kappa 11gn
19	10	100.0	1 N71032	Urate oxidase codl
20	10	100.0	1 N71032	Fragment D of urat
21	10	100.0	1 N71032	HSF CDNA sequence.
22	10	100.0	1 N71032	Sequence of thymos
23	10	100.0	1 N71032	Thymosin alpha 1 s
24	10	100.0	1 N71032	Sequence of synthe
25	10	100.0	1 N71032	Protein D - human
26	10	100.0	1 N71032	Newcastle Disease
27	10	100.0	1 N71032	DNA encoding envel
28	10	100.0	1 N71032	Clone CD20.4 encod
29	10	100.0	1 N71032	Human CD20 antigen
30	10	100.0	1 N71032	Sequence of plasmid
31	10	100.0	1 N71032	Sequence of oligos
32	10	100.0	1 N71032	Sequence of thymos
33	10	100.0	1 N71032	Adrenodoxin reduct
34	10	100.0	1 N71032	Sequence of synthe
35	10	100.0	1 N71032	Partial exotoxin A
36	10	100.0	1 N71032	Odotant receptor C
37	10	100.0	1 N71032	Korean hepatitis C
38	10	100.0	1 N71032	KHCV 937 probe P3
39	10	100.0	1 N71032	Expressed Sequence
40	10	100.0	1 N71032	Expressed Sequence
41	10	100.0	1 N71032	Expressed Sequence
42	10	100.0	1 N71032	Expressed Sequence
43	10	100.0	1 N71032	Expressed Sequence

ALIGNMENTS

ALIGNMENTS						
RESULT	1	10	100.0	163	1	039914
N90613	10	100.0	260	1	X27003	Expressed Sequence Differentially exp
ID	N90613 standard; cDNA; 1476 BP.					
AC	N90613;					
DE	20-DEC-1989 (first entry).					
CD	CD20 cDNA.					
KW	Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;					
KM	HIV box; immunoselection; immune deficiency diseases; vasculitis;					
PN	systemic lupus erythematosus; rheumatoid arthritis; neoplasms.					
PD	EP-330191-A.					
PF	30-AUG-1989.					
PR	23-FEB-1989; 103127.					
PA	25-FEB-1989; US-160416.					
PI	(GEO) Gen. Hospital Corp.					
PI	Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,					
PI	Simmons D, Stamenkovic I, Stengelin S;					
DR	WPI; 89-250302/35.					
PT	Rapid immunoselection cloning - used to clone genes encoding					
PT	cell surface antigens associated with mammalian T lymphocytes.					
PS	Disclosure; fig. 10; 69pp; English.					
CC	CD20.4 encodes a cell surface antigen involved in cell mediated immunity.					
CC	This DNA can be expressed in a vector which transforms COS cells. The					
CC	vector can isolate any protein, and clones are easy to manipulate.					
SQ	Sequence 1476 BP; 435 A; 324 C; 285 G; 432 T;					
QY	1 TCTCCTTAG 10					
DB	981 TCTCCTTAG 990					
DE	2					
ID	N90610 standard; cDNA; 1473 BP.					
AC	N90610;					
DE	20-DEC-1989 (first entry).					
CD	CD20.4 cDNA.					
KW	Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;					
KM	HIV box; immunoselection; immune deficiency diseases; vasculitis;					
KW	systemic lupus erythematosus; rheumatoid arthritis; neoplasms.					
FN	Key					
FT	cds					
FT	94.987					
FT	/*tag-a					
FT	1042					
FT	/*tag-b					
FT	/note= site of polyA+ tail in CD20.6					
PN	EP-330191-A.					
PD	30-AUG-1989.					
PF	23-FEB-1989; 103127.					
PR	25-FEB-1989; US-160416.					
PA	(GEO) Gen. Hospital Corp.					
PI	Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,					
PI	Simmons D, Stamenkovic I, Stengelin S;					
DR	WPI; 89-250302/35.					
P-PSDS	P91356.					

Query Match 100.0%; Score 10; DB 1; Length 1473;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10
 Db 979 TCTCCTTAG 988

RESULT 3
 N90225/C
 ID N90225 standard; DNA; 1982 BP.
 AC N90225;
 DT 1-NOV-1989 (first entry)
 DE Malaria-specific Plasmodium falciparum protein.
 KM DNA: malaria; 41kd protein; Plasmodium falciparum; vaccine.
 OS Plasmodium falciparum
 Fu Key Location/Qualifiers
 cds 526..1611
 /*tag- a

PN EP-322712-A.
 PD 05-JUN-1989.
 PE 20-DEC-1988; 121299.
 PF 30-DEC-1987; DE-831351.
 PR (BEHW) Behringwerke.
 PI Knapp B, Hundt E, Enders B, Kupper H;
 DR WPI: 89-194071/27.
 P-PSDB: P90419.
 PT New antigenic proteins from Plasmodium falciparum - new encoding nucleic acid sequences and derived antibodies, useful in vaccines, diagnosis etc.
 PS Claim 1; page 23 and Table 18; 25pp; German.
 CC The DNA encodes a malaria-specific P. falciparum 41kd protein (see P90419) is useful in a protective vaccine, esp. against malaria. Produced antibodies are useful for passive immunisation, and antibodies, the protein and the DNA sequence are useful as diagnostic reagents. See also N90211-25 and P90403-19.
 CC Sequence 1982 BP; 764 A; 230 C; 287 G; 701 T;
 SQ

Query Match 100.0%; Score 10; DB 1; Length 1982;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10
 Db 1302 TCTCCTTAG 1293

RESULT 4
 N90732/C
 ID N90732 standard; DNA; 5376 BP.
 AC N90732;
 DT 18-JUN-1990 (first entry)
 DE Sequence of plasmid p8/3
 KM Plasmid p8/3; vector; Plasmodium falciparum merozoite antigen.
 FH Key Location/Qualifiers
 FT cds 115..1257
 /*tag- a
 /*note-"It codes for a 41kd polypeptide"
 FT signal_peptide 115..168
 /*tag- b
 FT mat_peptide 169..1257
 /*tag- c
 PN EP-309746-A.
 PD 05-APR-1989.
 PE 27-AUG-1988; 114016.
 PF 08-SEP-1987; CH-003486.
 PR (HOFF) Hoffmann-La Roche AG.
 PA Certa U;
 PI WPI: 89-101095/14.
 DR New peptide(s) -
 PT

PT contg. epitope(s) of Plasmodium falciparum merozoite antigen,
 PT for use in malaria vaccines
 PS Figure 11a-11e; 67pp; German.
 CC Plasmid p8/3 is used as an expression vector.
 SQ Sequence 5376 BP; 1592 A; 1152 C; 1213 G; 1419 T;

Query Match 100.0%; Score 10; DB 1; Length 5376;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10
 Db 945 TCTCCTTAG 936

RESULT 5
 N90733/C
 ID N90733 standard; DNA; 1599 BP.
 AC N90733;
 DT 18-JUN-1990 (first entry)
 DE Genomic DNA from Plasmodium falciparum KI isolates encoding the 41 kd merozoite antigen
 KM Plasmodium falciparum KI isolate; merozoite antigen; epitope; immunogen;
 KW anti-malaria vaccine; passive immunisation; malaria diagnosis.
 OS Plasmodium falciparum
 Fu Key Location/Qualifiers
 FT cds 346..1434
 /*tag- a
 FT conflict 619..622
 /*tag- b
 /*note-"Differs from CDNA from M25 isolate of P. falciparum."
 FT conflict 718..721
 /*tag- c
 /*note-"Differs from CDNA from M25 isolate of P. falciparum"
 FT conflict 949..952
 /*tag- d
 /*note-"Differs from CDNA from M25 isolate of P. falciparum"

PN EP-309746-A.
 PD 05-APR-1989.
 PE 27-AUG-1988; 114016.
 PF 08-SEP-1987; CH-003486.
 PR (HOFF) Hoffmann-La Roche AG.
 PA Certa U;
 PI WPI: 89-101095/14.
 DR P-PSDB: P93566.
 PT New peptide(s) -
 PT contg. epitope(s) of Plasmodium falciparum merozoite antigen,
 PT for use in malaria vaccines
 PS Figure 12a-12d; 67pp; German.
 CC The genomic DNA of isolate RO-33 Ghana was largely identical to this in 3 codons (see n92937). Plasmodium falciparum merozoite antigen can be coupled to an affinity peptide, or adsorbed or covalently bonded on a carrier. It, and its derivatives, are useful as immunogens in anti-malaria vaccines. Antibodies directed against it are useful for passive immunisation and diagnosis.
 CC Sequence 1599 BP; 619 A; 207 C; 259 G; 514 T;
 SQ

Query Match 100.0%; Score 10; DB 1; Length 1599;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTTAG 10
 Db 1122 TCTCCTTAG 1113

RESULT 6

N91650/C
ID N91650 standard; DNA; 1426 BP.
AC N91650;
DE 16-MAR-1990 (first entry)
DE DNA encoding Sn-glycerol-3-phosphate acyl transferase.
KW Chloroplast; Cucurbita moschata
FH Key Location/Qualifiers
FT cds 124..1314
FT /*tag= a
PN J01235594-A.
PD 20-SEP-1989.
PE 14-MAR-1988; 060079.
PR 14-MAR-1988; JP-060079.
PA (MURA) Murata N.
DR WPI; 89-319095/44
DR P-PSDB; P93107.
DR Glycerol-3-phosphate acyl transferase coding DNA chain - useful for changing nature of chloroplast membrane lipid phosphatidyl glycerol. Claim 1; page 343; 7pp; Japanese.
CC DNA is introduced into various kinds of plants to alter the nature of the chloroplast membrane lipid phosphatidyl glycerol. It may also be used as a probe to allow cloning of enzyme genes which exist in closely related CC plants but which have slightly different substrate selectivity.
CC Sequence 1426 BP; 384 A; 322 C; 312 G; 408 T;
SQ

Query Match 100.0%; Score 10; DB 1; Length 1426;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAAG 10
Db 853 TCTCCTTAAG 844
|||||

RESULT 7
N92937/C
ID N92937 standard; DNA; 1254 BP.
AC N92937;
DE 18-JUN-1990 (first entry)
DE cDNA from Plasmodium falciparum M25 isolates encoding the 41 KD DE merozoite antigen
KW Plasmodium falciparum M25 isolate; merozoite antigen; epitope; immunogen;
KW anti-malaria vaccine; passive immunisation; malaria diagnosis.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 37..1125
FT /*tag= a
FT 640..643
FT /*tag= b
FT /note="Differs from genomic DNA in K1 isolate of P. falciparum"
FT 310..31
FT conflict
FT /*tag= c
FT /note="Differs from genomic DNA from K1 isolates of P. falciparum"
FT 407..410
FT /*tag= d
FT /note="Differs from genomic DNA in K1 isolate of P. falciparum"
PN EP-309746-A.
PD 05-APR-1989.
PE 27-AUG-1988; 114016.
PR 08-SEP-1987; CH-003486.
PA (HOFF) Hoffmann-La Roche AG.
PI Certa U;
DR WPI; 89-101095/14.
DR P-PSDB; P93567.
DR New peptide(s) -
PT contg. epitope(s) of Plasmodium falciparum merozoite antigen,
PT for use in malaria vaccines
PS Figure 12a-12d; 6/PP; German.
CC It differs from genomic DNA in K1 isolates of P. falciparum at 3 codons

CC (see n90733). Plasmodium falciparum merozoite antigen can be CC coupled to an affinity peptide, or adsorbed or covalently CC bonded on a carrier. It, and its derivatives, are useful as immunogens in CC anti-malaria vaccines. Antibodies directed against it are useful for CC passive immunisation and diagnosis.
SQ Sequence 1254 BP; 471 A; 186 C; 229 G; 368 T;

Query Match 100.0%; Score 10; DB 1; Length 1254;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAAG 10
Db 813 TCTCCTTAAG 804
|||||

RESULT 8
Q04266/C
ID Q04266 standard; cDNA; 1717 BP.
AC Q04266;
DE 13-SEP-1990 (first entry)
DE Probe containing coding exons for murine Ig light chain J regions
KW chimeric antibodies; TAG72; variable region of immunoglobulin light chain;
KW variable region of immunoglobulin heavy chain; light chain J region.
OS synthetic.
PN EP-365997-A.
PD 02-MAY-1990.
PE 18-OCT-1989; 119361.
PR 19-OCT-1989; US-259943.
PA (DOMC) Dow Chemical Co.
PI Mezes P, Gourille B, Rixon M;
DR WPI; 90-133521/18.
DR Chimeric antibodies against TAG72 and conjugate to provide imaging PT markers and therapeutic tools
PS Disclosure; PP; English.
CC The sequence is the 1.71kb HindIII-PstI fragment isolated from plasmid CC pBD1. It contains the coding exons for murine light chain J regions 1.e. CC J1-J5. The probe was used to isolate the C49 light chain for inclusion CC in chimeric antibodies. A typical chimera is composed of CC variable regions of heavy and light chains with high affinity for TAG72, CC (a tumour-associated glycoprotein) and constant regions from human CC antibodies. It is thought that these chimeric antibodies have fewer side- CC effects when administered to human patients due to the constant regions CC being of human origin.
CC See also Q04259-Q04265 and Q04267-9.
SQ Sequence 1717 BP; 492 A; 297 C; 425 G; 503 T;

Query Match 100.0%; Score 10; DB 1; Length 1717;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAAG 10
Db 554 TCTCCTTAAG 545
|||||

RESULT 9
N81290/C
ID N81290 standard; cDNA; 1764 BP.
AC N81290;
DE 22-OCT-1990 (first entry)
DE New recombinant cDNA coding for the fusion protein F0 of Newcastle DE disease virus (NDV)
KW Vaccines; poultry; clone; IL14(F); Newcastle disease virus; protein F0;
KW fowl vaccine; ss.
OS Newcastle disease virus.
FH Key Location/Qualifiers
FT signal_peptide 68..121
FT /*tag= a
FT 68..361
FT /*tag= b
FT cds


```

FT      cds      /product=Peptide F2
FT      362. 376
FT      /tag= C
FT      /product=Linking peptide
FT      377. 1648
FT      /tag= d
FT      /product=Peptide F1
PN      EP-252060-A.
PD      07-JAN-1988.
PF      29-JUN-1987. 870091.
PR      30-JUN-1986; US-880371.
PA      (SMIX) Smith Kline Rit SA.
PI      Espion D, Burny A, Debouck C;
DR      WPI: 88-001887/01.
DR      P-PSDB; P80986.
PT      Recombinant DNA encoding fusion protein of Newcastle disease virus -
PT      or its immunogenic polypeptide fragments, useful in poultry
PT      vaccine
PS      Claim 2: Pages 7-9 and Fig 1(a)-1(b); 21pp; French.
CC      A DNA molecule with this sequence is claimed. It was obtained as follows.
CC      Poly(A)-mRNA is isolated from NDV-infected BHK-21 cells used to construct
CC      complementary DNA by usual methods and this inserted into psti-cleaved
CC      pBR322. Recombinant plasmids were cloned in E.coli MM294 and selected by
CC      hybridisation with a viral RNA probe. Clone (designated II.14(F) contg.
CC      the longest insert was sequenced: it contains the claimed 1764-base
CC      sequence which codes for 526 AA protein consisting of two polypeptides
CC      linked by a pentapeptide region where proteolytic cleavage occurs. NDV
CC      fusion protein, and immunogenic polypeptide, are useful in vaccines for
CC      protecting poultry against NDV infection, and can be used diagnostically.
SQ      Sequence 1764 BP: 522 A; 415 C; 385 G; 442 T;

Query Match      100.0%; Score 10; DB 1; Length 1764;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCTCCTTAG 10
      111111111
Db      497 TCTCCTTAG 488

RESULT 10
ID      005796
ID      005796 standard; DNA; 1791 BP.
AC      005796;
DT      14-JAN-1991 (first entry)
DE      Encodes B.t.t insecticidal protein.
DE      B.t.t insecticidal protein; polyadenylation signals; ss.
KW      Bacillus thuringiensis tenebrionis.
EP      EP-385962-A.
PD      05-SEP-1990.
PF      20-FEB-1990; 870025.
PR      24-FEB-1989; US-315355.
PR      12-FEB-1990; US-476661.
PA      (MONS ) MONSANTO CO.
PI      Fischhoff DA, Perlak FJ;
DR      WPI: 90-269905/36.
PT      Increasing expression of heterologous genes in plants - by
PT      removing poly:adenylation signals in the gene.
PS      Claim 22; Page 77; 14pp; English.
CC      The insecticidal protein gene of B.t.t. was modified by site-
CC      directed mutagenesis to remove all putative polyadenylation signals
CC      and ATTA sequences. The mutated sequence was inserted into a plant
CC      transformation vector such as pMON93. The vector is then inserted
CC      into an Agrobacterium strain and used to transform potato, tomato
CC      and tobacco plants. Transformed plants were found to be resistant
CC      to Colorado Potato Beetle.
CC      See also 005790-005795, 005797-005800 and 006561.
SQ      Sequence 1791 BP: 501 A; 467 C; 391 G; 432 T;

Query Match      100.0%; Score 10; DB 1; Length 1791;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;

```

```

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCTCCTTAG 10
      111111111
Db      477 TCTCCTTAG 486

RESULT 11
ID      N71032/C
ID      N71032 standard; DNA; 102 BP.
AC      N71032;
DT      17-APR-1991 (first entry)
DE      Sequence of synthetic thymosin Alpha I gene.
KW      Hormone; expression vector; ds.
FH      Key
FT      cds      Location/Qualifiers
FT      6. 95
FT      /tag= a
FT      1. 4
FT      misc-feature
FT      /tag= b
FT      /label= sticky end
FT      misc-feature
FT      complement (99. 102)
FT      /tag= c
FT      /label= sticky end
PN      US4650761-A.
PD      17-MAR-1987.
PF      09-NOV-1983; 550167.
PR      27-NOV-1981; US-325511.
PR      09-NOV-1983; US-550167.
PA      (ELIL ) Eli Lilly & CO.
PI      Hershenberger CL, Rostock PR;
DR      WPI: 87-093428/13.
PT      P-PSDB; P70673.
PT      Improved stabilisation and selection of E coli transformants -
PT      using DNA vector contg. PstI- Hinc II CI repressor and gene
PT      encoding human pro-insulin
PS      Example: Fig 7; 24pp; English.
CC      Plasmid pin alpha 1, which is based on pBR322, contains a gene
CC      specifying ampicillin resistance and a structural gene specifying
CC      thymosin alpha 1 cloned at its 5' coding strand end into an EcoRI
CC      site at its 3' end into a BamHI site.
SQ      Sequence 102 BP: 33 A; 14 C; 25 G; 30 T;

Query Match      100.0%; Score 10; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCTCCTTAG 10
      111111111
Db      63 TCTCCTTAG 54

RESULT 12
ID      010157
ID      010157 standard; DNA; 3744 BP.
AC      010157;
DT      15-MAR-1991 (first entry)
DE      Sequence contg. abaa gene of Aspergillus nidulans.
KW      abaa gene; inducible promoter; filamentous fungi growth; ss.
OS      Aspergillus nidulans.
FH      Key
FT      exon      Location/Qualifiers
FT      883. 1333
FT      /tag= a
FT      /label= exon I of abaa gene
FT      1384. 1560
FT      /tag= b
FT      /label= exon II of abaa gene
FT      1612. 3374
FT      /tag= c
FT      /label= exon III of abaa gene
PN      M09015144-A.
PD      13-DEC-1990.
PF      01-JUN-1990; U03125.

```


PR 02-JUN-1989; US-360294.
 PA (UYGE-) UNIV GEORGIA RES FO.
 PI Timberlake W, Adams TE, Miralito PM;
 DR WPI: 91-007214/01.
 DE R-PSDB; R10047.
 CC DNA sequence of abaa gene of *Aspergillus nidulans* and gene fusion
 PT - with inducible promoter for control of growth and
 PS differentiation of filamentous fungi
 CC This DNA fragment encodes the A.nidulans abacus (abaa) gene follow-
 ing processing to remove the 2 introns. The protein encoded is
 CC capable of regulating growth and development in filamentous fungi.
 CC It is pref. fused to an inducible promoter, e.g. nad or prnA,
 CC enabling the mis-scheduled expression of the abaa gene, inducing
 CC abortive development in the fungus. This enables increased prodn.
 CC of medically or commercially important substances which are normally
 CC only produced by the fungi during conidiation.
 Sequence 3744 BP; 933 A; 1100 C; 793 G; 918 T;

Query Match 100.0%; Score 10; DB 1; Length 3744;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTAAG 10
 |||||
 DB 176 TCTCCTAAG 185

RESULT 13
 Q10869/c
 ID Q10869 standard; DNA; 1470 BP.
 AC Q10869;
 DT 15-APR-1991 (first entry)
 DE Plasmodium vivax 66KD merozoite antigen gene partial sequence.
 KW Plasmodium merozoite antigens; malaria vaccine; ss.
 OS Plasmodium vivax.
 FH Key
 FT cds
 FT 1.1470
 FT Location/Qualifiers
 FT /*tag= a
 FT /*product= part of P.vivax merozoite antigen protein
 PD US7483516-A.
 PD 13-JAN-1991;
 PD 22-FEB-1990; 483516.
 PR (USSH) NAT INST OF HEALTH.
 PR Waters AP, McCutchan TF;
 WPI: 91-044381/06.
 DR P-PSDB; R10935.
 PT Plasmodium merozoite antigen proteins and DNA sequences - useful
 PT in vaccine prodn., anti-malarial drug design, and in diagnostics
 PS Disclosure: fig 1; 31pp; English.
 CC This sequence is contained in a recombinant DNA molecule com-
 CC prising e.g. a pUC19 or vaccinia virus vector sequence and
 CC regulatory elements. This allows efficient expression of the
 CC antigen on transformation of host cells. The P.vivax antigen
 CC prodn. is useful in an anti-malarial vaccine for conferring immu-
 CC nity against the merozoite form of the parasite and preventing in-
 CC fection of uninfected red blood cells. See also Q10418.
 Sequence 1470 BP; 496 A; 279 C; 361 G; 334 T;

Query Match 100.0%; Score 10; DB 1; Length 1470;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTAAG 10
 |||||
 DB 214 TCTCCTAAG 205

RESULT 14
 Q10245/c

ID Q10245 standard; cDNA; 1121 BP.
 AC Q10245;
 DT 27-MAR-1991 (first entry)
 DE Clone 9C encoding a protein with urate oxidase activity.
 KW urate oxidase; uricase; hyperuricaemia; kidney stones; chemotherapy;
 SS.
 OS *Aspergillus flavus*.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 109.1017
 FT /*tag= a
 FT /*product= urate oxidase-like protein

EP-408461-A.
 PD 16-JAN-1991.
 PD 13-JUL-1990; 402023.
 PR 13-JUL-1989; FR-009550.
 PR 29-DEC-1989; FR-017466.
 PR 06-FEB-1990; FR-001368.
 PA (SNFI) SANOFI SA.
 PI Caput D, Ferrara P, Guillemot JC, Kaghad M, Legoux R;
 PI Loison G, Labree, Lupker J;
 DR WPI: 91-016644/03.
 DE P-PSDB; R10222.
 PT New protein with high urate oxidase activity - and recombinant
 PT DNA encoding it, vectors and transformed cells, used for treating
 PT hyperuricaemia, etc.
 PS Claim 8; Fig 3; 68pp; French.
 CC mRNA was isolated from a urate oxidase-producing strain of *A.flavus*
 CC and reverse transcribed into cDNA. A gene bank was constructed from
 CC E.coli MC 1061 containing plasmid pTZ19R containing the cDNA. Probes
 CC were synthesised based upon the amino acid sequence of A.flavus
 CC urate oxidase and these used to screen the gene bank. Five positive
 CC colonies were analysed; this is the sequence of the longest cDNA
 CC insert ("clone 9C"). See also Q10246-Q10253, Q10253.
 Sequence 1121 BP; 281 A; 333 C; 264 G; 243 T;

Query Match 100.0%; Score 10; DB 1; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCTCCTAAG 10
 |||||
 DB 628 TCTCCTAAG 619

RESULT 15
 Q10246/c
 ID Q10246 standard; DNA; 906 BP.
 AC Q10246;
 DT 27-MAR-1991 (first entry)
 DE Encodes urate oxidase specifically for expression in prokaryotes.
 KW urate oxidase; uricase; hyperuricaemia; kidney stones; chemotherapy;
 SS.
 OS Synthetic.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 1.906
 FT /*tag= a
 FT /*product= urate oxidase

EP-408461-A.
 PD 16-JAN-1991.
 PD 13-JUL-1990; 402023.
 PR 13-JUL-1989; FR-009550.
 PR 29-DEC-1989; FR-017466.
 PR 06-FEB-1990; FR-001368.
 PA (SNFI) SANOFI SA.
 PI Caput D, Ferrara P, Guillemot JC, Kaghad M, Legoux R;
 PI Loison G, Labree, Lupker J;
 DR WPI: 91-016644/03.
 DE P-PSDB; R10222.
 PT New protein with high urate oxidase activity - and recombinant
 PT DNA encoding it, vectors and transformed cells, used for treating
 PT hyperuricaemia, etc.
 PS Claim 10; Page 50; 68pp; French.

CC mRNA was isolated from a urate oxidase-producing strain of A.flavus
CC and reverse transcribed into cDNA. A gene bank was constructed from
CC E.coli MC 1061 containing plasmid pTZ19 containing the cDNA. Probes
CC were synthesised based upon the amino acid sequence of A.flavus
CC urate oxidase and these used to screen the gene bank. Five positive
CC clones were analysed, one of which contained inserts of 1.2kb (-
CC "clone 9c"). An AccI-XbaI fragment was isolated from 9c and ligated
CC to a synthetic fragment designed to replace codons immediately
CC downstream of the start ATG with codons typical of prokaryotic
CC genes (substituted bases are found at positions 6, 9, 21 and 42).
CC The recombinant sequence was ligated to plasmid p462 to give p466.
CC See also Q10245 and Q10247-Q10253 and Q10255.
SQ Sequence 906 BP; 219 A; 267 C; 241 G; 179 T;

Query Match 100.0%; Score 10; DB 1; Length 906;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0" 1 TCTCCTTAG 10
|||||||
520 TCTCCTTAG 511

Search completed: September 18, 1999, 00:33:28
Job time: 18968 sec


```

FT cds 405.1838
/*tag- a
PN ZA8809415-A.
PD 27-SEP-1989.
PF 15-DEC-1988; 889415.
PR 18-DEC-1987; US-124981.
PA (CHIR) Chiron Corp.
PI Valenzuela P, Brake A, Randolph A;
DR WPI: 90-007677/01.
PT Crocalidus venom fibrolyase prodn.
PS also yeast expression vectors contg. DNA and recombinant DNA.
CC Example 5; Page 29; 64pp; English.
CC PK3311 plasmid as carried by the pAB24 yeast vector with GAP promoter.
CC Fibrolyases are thrombolytic agents, useful in treating thromboembolic
CC conditions e.g. pulmonary embolism.
CC See also Q02837 and Q02838.
SQ Sequence 2809 BP; 887 A; 552 C; 538 G; 832 T;

Query Match 12.7%; Score 7; DB 1; Length 2809;
Best Local Similarity 100.0%; Pred. No. 7.6e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
DB 2286 CTGCACG 2280

RESULT 3
002894
ID Q02894 standard; DNA: 2026 BP.
AC Q02894;
DT 11-JUL-1990 (first entry)
DE cDNA insert of vector PSP 6K-13-7 encoding pulmonary surfactant protein.
KW Respiratory distress syndrome; RDS; hyaline membrane
KW disease; pulmonary surfactant protein.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 14.1156
FT /*tag- a
FT mat_peptide 614.811
FT /*tag- b

PN US482422-A.
PD 21-NOV-1989.
PF 24-SEP-1987; 100372.
PR 26-SEP-1985; US-791120.
PR 15-AUG-1986; US-897183.
PR (GENE-) Genetics Institute.
PR Taenush HW, Jacobs RA, Steinbrink DR, Floros J, Phelps DS, Fritsch EF;
PR WPI: 90-036829/05.
DR P-PSDB: R05093.
PT Purified human pulmonary surfactant protein -
PT useful for treating respiratory distress syndrome.
PS Disclosure: Table 6; 15pp; English.
CC Protein is useful in treatment of respiratory distress syndrome (Hyaline
CC membrane disease) enhancing pulmonary surfactant activity.
CC Protein is encoded by the cDNA insert in vector PSP 6K-13-7.
SQ Sequence 2026 BP; 442 A; 626 C; 532 G; 425 T; 1 Others;

Query Match 12.7%; Score 7; DB 1; Length 2026;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
DB 1722 CTGCACG 1728

RESULT 4
003206/c
ID Q03206 standard; DNA: 117 BP.

```

```

AC Q03206;
DT 10-JUL-1990 (first entry)
DE Gene encoding protein associated with ras oncogene.
KW ras oncogene; cancer; GTPase; ds.
OS Synthetic.
PN WO9000179-A.
PD 11-JAN-1990.
PF 4-JUL-1989; F00348.
PR 4-JUL-1988; FR-009031.
PA (INRW) Inserm Inst Nat Sante.
PI Tavittian A, Pizon V, Chardin P;
DR WPI: 90-037122/05.
DR P-PSDB: R05070.
PT DNA encoding sequences associated with human oncogenes -
PT and derived antibodies, useful for in vitro diagnosis of cancer
PT Claim 8; Page 59; 92pp; French.
PS Claim 8; Page 59; 92pp; French.
CC Derived peptides can fix GTP and GDP and have GTPase activity. Abs raised
CC to these peptides can diagnose associated diseases - absence or excess of
CC the product indicates loss of balance between anti- and oncogenic factors
CC in the cells. The product may be used to treat conditions associated with
CC ras or rap genes.
SQ Sequence 117 BP; 24 A; 36 C; 34 G; 23 T;

Query Match 12.7%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 8.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
DB 15 CTGCACG 9

RESULT 5
003212/c
ID Q03212 standard; DNA: 558 BP.
AC Q03212;
DT 10-JUL-1990 (first entry)
DE RAP2 Gene encoding at least one peptide associated with ras oncogene.
KW ras oncogene; cancer; GTPase; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 4.552
FT /*tag- a

PN WO9000179-A.
PD 11-JAN-1990.
PF 4-JUL-1989; F00348.
PR 4-JUL-1988; FR-009031.
PA (INRW) Inserm Inst Nat Sante.
PI Tavittian A, Pizon V, Chardin P;
DR WPI: 90-037122/05.
DR P-PSDB: R05076.
PT DNA encoding sequences associated with human oncogenes -
PT and derived antibodies, useful for in vitro diagnosis of cancer
PT Claim 8; Page 59; 92pp; French.
PS Disclosure: Fig 3a-b; 92pp; French.
CC Derived peptides can fix GTP and GDP and have GTPase activity. Abs raised
CC to these peptides can diagnose associated diseases - absence or excess of
CC the product indicates loss of balance between anti- and oncogenic factors
CC in the cells. The product may be used to treat conditions associated with
CC ras or rap genes.
SQ Sequence 558 BP; 142 A; 143 C; 165 G; 108 T;

Query Match 12.7%; Score 7; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
DB 69 CTGCACG 63

```



```

RESULT 6
ID 003247/c standard; DNA; 1766 BP.
AC 003247;
DE 11-JUL-1990 (first entry)
DE Aspergillopepsin A gene.
KM Aspartic proteinase; chymosin; t-PA; human serum albumin ds.
OS Aspergillus awamori.
FH Key Location/Qualifiers
FT signal_peptide 178..489
FT FT /tag- a
FT FT 238..385
FT FT precursor_rna
FT FT /tag- b
FT FT /product-Highly charged propeptide.
FT FT cds 178..1521
FT FT /tag- c
FT FT mat_peptide 386..1521
FT FT /tag- d
FT FT intron 490..540
FT FT /tag- e
FT FT intron 827..878
FT FT /tag- f
FT FT intron 1128..1185
FT FT /tag- g
FT FT poly_a_site 1573..1577
FT FT /tag- h
FT FT /evidence=Putative.
PN W09000192-A.
PD 11-JAN-1990.
PF 01-JUL-1989; U02891.
PR 01-JUL-1988; US-214237.
PA (GENE-) Genencor Inc.
PI Berka R, Hayenga K, Lawlis VB, Ward M;
DR WPI: 90-037128/05.
DR P-PSDB: R05078.
PT Aspartic proteinase deficient filamentous fungi prepn.
PT by pref. using a gene replacement vector, useful as hosts for
PT increased prodn. of heterologous polypeptide(s).
PS Disclosure; Fig 3; 64pp; English.
CC Mutating the aspergillopepsin gene, a strain deficient in the aspartic
CC proteinase may be synthesised. The absence of such proteolytic enzymes
CC improves the production of heterologous gene products, for instance
CC bovine chymosin, t-PA; human serum albumin or mammalian growth hormone
CC may be produced.
CC Sequence 1766 BP; 388 A; 529 C; 430 G; 419 T;

Query Match 12.7%; Score 7; DB 1; Length 1766;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
DB 951 CTCGACG 945

RESULT 7
ID 003216 standard; DNA; 2122 BP.
AC 003216;
DE 11-JUL-1990 (first entry)
DE SalI restriction fragment of plasmid pUC-RNSII containing rabies
DE preglycoprotein G protein.
KM Rabies; pUC-RNSII; pre-glycoprotein G; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 63..1635
FT FT /tag- a
PN W09000191-A.
PD 11-JAN-1990.
PF 27-JUN-1989; U02809.
PR 01-JUL-1988; US-214085.

```

```

PA (SALK) Salk Inst for Biol Stud.
PI Tschopp JF, Dixon DB;
DR WPI: 90-037127/05.
DR P-PSDB: R05078.
PT Prodn. of rabies glyco-protein G
PT by culturing P.pastoris contg. DNA encoding preprotein with
PT methanol as carbon source.
PS Disclosure; Table 1; 82pp; English.
CC Isolating rabies preglycoprotein G, and associating it with a promoter
CC segment of a methyloctrophic yeast strain and a terminator from a second
CC methyloctrophic strain, allows the protein to be selectively expressed in
CC P. pastoris culture (pref. ATCC 20880).
CC The inactive virus protein is useful in a vaccine as a prophylactic
CC measure for at risk individuals and animals.
CC Sequence 2122 BP; 634 A; 458 C; 487 G; 543 T;

Query Match 12.7%; Score 7; DB 1; Length 2122;
Best Local Similarity 100.0%; Pred. No. 7.7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
DB 899 CTCGACG 905

RESULT 8
ID 003246/c standard; DNA; 3168 BP.
AC 003246;
DE 12-JUL-1990 (first entry)
DE PelC Pectin lyase PLC recombinant DNA molecule.
KM Pectin lyase; PLC; expression system; interferon.
OS Aspergillus niger.
FH Key Location/Qualifiers
FT cds 1368..2707.
FT FT /tag- a
FT FT intron 1854..1931
FT FT /tag- b
FT FT intron 1998..2062
FT FT /tag- c
FT FT intron 2234..2296
FT FT /tag- d
PN EP-353188-A.
PD 31-JAN-1990.
PF 19-JUL-1989.
PR 28-JUL-1988; GB-018046.
PR 26-JUN-1989; GB-014666.
PA (CIBA) Ciba Geigy AG.
PI Helm J, Meyhack B, Visser J;
DR WPI: 90-031714/05.
DR P-PSDB: R05112.
PT DNA coding for pectin lyase(s) PLA, PLB, PLC, PLF or PLF -
PT used to produce pectin lyase(s) in pure form for constructing
PT hybrid vectors expressing foreign genes.
PS Claim 5; Fig 12; 63pp; English.
CC Useful in production of hybrid vectors expressing the pectin lyase
CC gene and/or hybrid vectors expressing foreign genes eg. interferon
CC within a filamentous fungal expression system esp. Aspergillus niger.
CC Sequence 3168 BP; 823 A; 780 C; 720 G; 845 T;

Query Match 12.7%; Score 7; DB 1; Length 3168;
Best Local Similarity 100.0%; Pred. No. 7.6e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTCGACG 55
DB 1823 CTCGACG 1817

RESULT 9

```


003245
ID 003245 standard; DNA; 2774 BP.
AC 003245;
DT 12-JUL-1990 (first entry)
DE PclB Pectin lyase PUB recombinant DNA molecule.
KW Pectin lyase; PUB; expression system; interferon.
OS Aspergillus niger.
FH Key Location/Qualifiers
FT CDS 1134..2502
FT /tag= a
FT Intron 1338..1400
FT /tag= b
FT Intron 1544..1599
FT /tag= c
FT Intron 1726..1782
FT /tag= d
FT Intron 2113..2169
FT /tag= e
PN EP-353188-A.
31-JAN-1990.
19-JUL-1989.
PK 28-JUL-1988; GB-018046.
PR 26-JUN-1989; GB-014666.
PA (CIBA) Ciba Geigy AG.
PI Heim J, Meyhack B, Visser J;
DR WPI; 90-031714/05.
DR P-PSDB; R05111.
PT DNA coding for pectin lyase(s) PLA, PLB, PLC, PLE or PLF -
PT used to produce pectin lyase(s) in pure form for constructing
PT hybrid vectors expressing foreign genes.
PS Claim 4; Fig 11: 63pp; English.
CC Useful in production of hybrid vectors expressing the pectin lyase
CC gene and/or hybrid vectors expressing foreign genes eg. interferon
CC within a filamentous fungal expression system esp. Aspergillus niger.
SQ Sequence 2774 BP; 653 A; 785 C; 673 G; 663 T;

Query Match 12.7%; Score 7; DB 1; Length 2774;
Best Local Similarity 100.0%; Pred. No. 7.7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
|||||
Db 2104 CTGCACG 2110

RESULT 10
232
003232 standard; RNA; 1763 BP.
DT 12-JUL-1990 (first entry)
DE Gene encodes function of maintaining a plasmid stably in a coryne-form
DE bacteria of the genus Brevibacterium.
KW PBV503; tryptophan synthetase; coryneform; ds.
OS Brevibacterium stationis.
PN EP-352763-A.
31-JAN-1990.
PF 26-JUL-1989; 113775.
PR 27-JUL-1988; JP-185428.
PR 08-SEP-1988; JP-223399.
PR 25-JAN-1989; JP-014098.
PA (MITS) Mitsubishi Petroch KK.
PI Kohama K, Kobayashi M, Kurusu Y, Yukawa H;
DR WPI; 90-03147/05.
PT New DNA fragments which stabilise plasmids in coryneform bacteria -
PT and derived recombinant vectors which are retained in cells
PT without selection pressure.
PS Claim 6; Page 23: 27pp; English.
CC Plasmid PBV503 contains DNA fragment present in Brevibacterium stationis
CC rfo 12144, contains gene which stabilises plasmids in the bacteria,
CC allowing maintenance with no selection pressure.
CC Tryptophan synthetase gene with associated promoter/operator system has

CC been maintained within a stable plasmid associated with the gene.
SQ Sequence 1763 BP; 433 A; 436 C; 445 G; 449 T;

Query Match 12.7%; Score 7; DB 1; Length 1763;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
|||||
Db 1625 CTGCACG 1631

RESULT 11
003235
ID 003235 standard; DNA; 354 BP.
AC 003235;
DT 12-JUL-1990 (first entry)
DE PB123 secretion expression plasmid for prodn. of human epidermal growth
DE factor (heGF).
KW Human epidermal growth factor; heGF; anti-cancer agent;
KW cancer; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT Promoter 6..91
FT /tag= a
FT /label= lrp.
FT signal_peptide 92..154
FT /tag= b
FT /label= E.coli alkaline phosphatase.
FT mat_peptide 155..313
FT /tag= c
FT /label= heGF gene.
FT terminator 327..350
FT /tag= d
FT /label= lrp A.

EP-352839-A.
31-JAN-1990.
PF 7-JUL-1989; 201822.
PR 15-JUL-1988; JP-177685.
PA (NNSH) Nippon Shinyaku KK.
PI Yano J, Murai M;
DR WPI; 90-031419/05.
PT Prodn. of human epidermal growth factor -
PT using Escherichia coli system including tryptophan promoter and
PT E.coli alkaline phosphatase signal peptide gene.
PS Disclosure; Fig 4; 23pp; English.
CC An E.coli strain transformed with the vector is useful for producing the
CC heGF in large pure quantities, with glycerol and/or glucose as a C-source
CC The heGF is useful as an anti-cancer agent.
SQ Sequence 354 BP; 97 A; 73 C; 82 G; 102 T;

Query Match 12.7%; Score 7; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 CTGCACG 55
|||||
Db 197 CTGCACG 203

RESULT 12
002814
ID 002814 standard; cDNA; 2537 BP.
AC 002814;
DT 10-MAR-1993 (revised)
DT 31-MAY-1989 (first entry)
DE Sequence of pre-TGF-beta1 cDNA.
KW Transforming growth factor beta-3 (TGF beta 3); tumour cells; growth
KW inhibition.
FH Key Location/Qualifiers
FT cds 842..2011


```

FT      /*tag= a
FT      /label-pre-TGF beta 1
FT      1677..2011
FT      cds
FT      /*tag= b
FT      /label-mature TGF-beta 1
FT      2015..2092
FT      gC-signal
FT      /*tag= c
FT      2093..2099
FT      misc-feature
FT      /*tag= d
FT      /label-TARA-like sequence
FT      37..113
FT      stem_loop
FT      /*tag= e
FT      863..911
FT      misc-feature
FT      /*tag= f
FT      /label-hydrophobic domain
PN      W08912101.A.
PD      14-DEC-1989.
        08-JUN-1988; WO-001945.
        08-JUN-1988; WO-001945.
        (GETH) Genentech Inc.
PI      Dernyck RM, Goeddel DV;
        WPI: 90-007474/01.
DR      P-PSDB: R04034.
PT      Nucleotide sequence encoding transforming growth factor beta-3 -used as a
        probe, or to produce TGF beta 3, for growth inhibition of certain normal
        PT and neoplastic cells, eg A549.
PS      Disclosure; Fig. 1b; 61pp; English.
CC      Sequence encodes the 380 amino acid (AA) precursor transforming growth
        factor-beta 1 (pre-TGF-beta 1) polypeptide. The 5' untranslated region of
        the TGF-beta 1 mRNA is 841 bases long, is purine rich and has a region of
        potential secondary structure. The TARA-like sequence in the 3' untrans-
        lated region of the gene is presumably a polyadenylation signal. Mature
        TGF-beta 1 comprises the C-terminal 112 AA's of pre-TGF-beta 1 and is
        cleaved at the Arg-Arg dipeptide preceding its NH2 terminus. The nucleic
        acid encoding the second subtype of TGF-beta (TGF-beta 3) is useful as a
        probe or to produce TGF-beta 3 for inhibition of growth of normal and
        neoplastic cells.
SQ      Sequence 2537 BP; 473 A; 893 C; 739 G; 432 T;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 2537;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        493 CTGCACG 499

RESULT 13
ID      002819 standard; DNA; 2676 BP.
AC      002819;
DT      10-MAR-1993 (revised)
DT      31-MAY-1989 (first entry)
DE      cDNA sequence encoding porcine TGF-beta 3.
KW      Transforming growth factor beta-3 (TGF beta 3); tumour cells; growth
        inhibition.
PN      W08912101.A.
PD      14-DEC-1989.
        08-JUN-1988; WO-001945.
        08-JUN-1988; WO-001945.
        (GETH) Genentech Inc.
PI      Dernyck RM, Goeddel DV;
        WPI: 90-007474/01.
DR      P-PSDB: R04080.
PT      Nucleotide sequence encoding transforming growth factor beta-3 -used as a
        probe, or to produce TGF beta 3, for inhibition of growth of normal
        PT and neoplastic cells, eg A549.
PS      Disclosure; Fig. 4; 61pp; English.
CC      This sequence encodes porcine transforming growth factor-beta 3 (TGF-
        beta 3) polypeptide. The nucleic acid sequence encoding this subtype is
        useful as a probe or to produce TGF-beta 3 for both normal and neoplastic

```

```

CC      cell growth inhibition.
SQ      Sequence 2676 BP; 704 A; 705 C; 699 G; 568 T;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 2676;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        193 CTGCACG 199

Db      193 CTGCACG 199

RESULT 14
ID      003218 standard; DNA; 1389 BP.
AC      003218;
DT      10-MAR-1993 (revised)
DT      12-JUL-1989 (first entry)
DE      DNA sequence encoding human protein C mutant
        human protein C; anti-coagulant; ss.
KW      Homo sapiens.
OS      Homo sapiens.
FH      Key
FT      cds
FT      1..1383
FT      /*tag= a
FT      /label-Human protein C mutant
PN      EP-352651-A.
PD      31-JAN-1990.
        21-JUL-1989; 113414.
        26-JUL-1988; JP-184538.
        (FARH) Hoechst Japan Ltd.
PI      Hashimoto T, Sato M;
        WPI: 90-031241/05.
DR      P-PSDB: R04800.
PT      Human protein C mutant with residues 156 and 157 replaced or deleted -
        has reduced activation rate and extended duration of activity as
        PT anti-coagulant.
PS      Disclosure; Page 11-14; 22pp; English.
CC      This sequence encodes a protein which differs from human protein C
        at positions 156 and 157 (Lys-Arg is substituted by Asn-Ser). This
        CC creates a single uncleaved chain. The mutant product shows anti-coagulant
        CC activity when activated by the thrombin-thrombomodulin complex. It has
        CC a lower activation rate and longer duration of activity compared with
        CC the native protein.
SQ      Sequence 1389 BP; 289 A; 419 C; 433 G; 248 T;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 1389;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      49 CTGCACG 55
        |||||||
        438 CTGCACG 444

Db      438 CTGCACG 444

RESULT 15
ID      001794 standard; DNA; 1383 BP.
AC      001794;
DT      25-JUL-1990 (first entry)
DE      Nucleotide sequence coding for hybrid human/bovine protein C
        human protein C; bovine protein C; Gla domain; blood anticoagulant;
        KW blood coagulation disorders; blood fibrinolysis acceleration.
PN      EP-354504-A.
PD      14-FEB-1990.
        05-AUG-1989; 114505.
        09-AUG-1988; JP-197144.
        (FARH) Hoechst Japan Ltd.
PI      Hashimoto T, Takahashi M;
        WPI: 90-046218/07.
DR      P-PSDB: R05211.
PT      Human protein C deriv. -

```


PT having Gla domain replaced by the bovine Gla to increase calcium
 binding activity and improve activation
 PS Claim 3, Page 14, 17pp; English.
 CC Also claimed are its encoded human protein C, the gene structure contg.
 CC it and a host cell contg. the gene structure. The N-terminal region of
 CC human protein C having gamma-carboxylated glutamic acid residues (Gla
 CC domain) has been replaced by the Gla domain of bovine protein C or by an
 CC equivalent of this bovine protein sequence w.r.t. its calcium binding
 CC activity, and/or its enhanced protein C activity. PCS4 was constructed
 CC from PCS1. DNA coding for part of protein C was removed from PCS4 by
 CC Sall digestion at a site just upstream of the leader sequence-coding
 CC sequence and again at a site corresp. to the 45th (Val) and 46th (Asp)
 CC residues of protein C. A synthetic DNA fragment was inserted into PCS4
 CC between these Sall sites and introduced into E. coli K12/HR101. The
 CC transformants harbouring a plasmid in an appropriate orientation, named
 CC PCS8, were screened and cultivated. PHSG293 was also constructed which
 CC had a neo gene as a selective marker and BstXI site providing the same
 CC asymmetric cohesive ends as PCS8 by digestion. The PCS8 was digested
 CC with BstXI and the 2 kb fragment was ligated with BstXI-digested PHSG293.
 CC The ligated DNA was packaged in vitro in a lambda phage packaging mixt.
 CC and transfected into E. coli K12/Om206. The DNA was isolated and
 CC introduced into CHO cells which were cultured to produce the hybrid
 CC protein C. Activated protein C inhibits blood coagulation or accelerates
 CC blood fibrinolysis and is used for treating blood coagulation disorders.
 SQ Sequence 1383 BP; 288 A; 413 C; 442 G; 240 T;

Query Match 12.78; Score 7; DB 1; Length 1383;
 Best Local Similarity 100.0%; Pred. No. 7.8e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CTGCACG 55
 Db 432 CTGCACG 438

Search completed: September 18, 1999, 00:33:26
 Job time: 18966 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:27:52 : Search time 2825.05 Seconds
(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-85
Perfect score: 10
Sequence: 1 TCCTCTTAG 10

Scoring table: IDENTITY_NUC

arched: 2546578 seqs, 98626752 residues

-abase :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100.0	38	20	D15131	D15131	RICCO129A R
2	100.0	435	20	D15980	D15980	RICCO1808A R
3	100.0	405	20	D21672	D21672	MUS81D09 MO
4	100.0	384	20	D22924	D22924	RICCI194A R
5	100.0	347	20	D24072	D24072	RICRI057A R
6	100.0	430	20	D24444	D24444	RICRI032A R
7	100.0	318	20	D24993	D24993	RICR2912A R
8	100.0	205	20	D25754	D25754	HUMGS04121
9	100.0	310	20	D31273	D31273	HUMI13202 H
10	100.0	414	20	D34553	D34553	CELR0082R
11	100.0	315	20	D35356	D35356	CELR001DYF
12	100.0	360	20	D35439	D35439	CELR020E5F
13	100.0	330	20	D36191	D36191	CELR002GZF
14	100.0	444	20	D38872	D38872	RICCI180A R
15	100.0	407	20	D38092	D38092	RICRI16A R
16	100.0	344	20	L26626	L26626	MUSB045A Ia
17	100.0	252	20	L36675	L36675	BNASTF756
18	100.0	376	20	M62140	M62140	ESTP0201 H1
19	100.0	163	20	M62160	M62160	ESTP0222 H1
20	100.0	440	20	M62264	M62264	ESTP0333 H1
21	100.0	418	20	M79639	M79639	WEST00176 M
22	100.0	460	20	M79767	M79767	WEST00304 M
23	100.0	417	20	M79768	M79768	WEST00305 M
24	100.0	422	20	M79840	M79840	WEST00377 M
25	100.0	427	20	T00381	T00381	WEST01102 E
26	100.0	339	20	T00955	T00955	WEST01680 E
27	100.0	406	20	T00997	T00997	WEST01718 E
28	100.0	176	20	T01374	T01374	WEST02095 E
29	100.0	357	20	T01649	T01649	WEST02370 E
30	100.0	417	20	T02220	T02220	WEST02941 E
31	100.0	311	20	T05175	T05175	ESTP03063 Fe
32	100.0	351	20	T06527	T06527	ESTP04416 Fe
33	100.0	398	20	T06573	T06573	ESTP04462 Fe
34	100.0	374	20	T06602	T06602	ESTP04491 Fe
35	100.0	354	20	T08281	T08281	ESTP06172 In
36	100.0	393	20	T09065	T09065	ESTP06958 In
37	100.0	194	20	T10525	T10525	hbc706 Huma
38	100.0	372	20	T13618	T13618	1783 Lamda
39	100.0	235	20	T16466	T16466	NIB1349 Nor
40	100.0	330	20	T18018	T18018	0360C3 CBSP
41	100.0	579	20	T18334	T18334	GUP 8-3 all
42	100.0	202	20	T20053	T20053	B360F Heart
43	100.0	349	20	T24010	T24010	seg2166 3HF
44	100.0	455	20	T24372	T24372	CRS1502 lam
45	100.0	317	51	AT15497	AT15497	UI-R-YO-a

ALIGNMENTS

RESULT	1	D15131	328 bp	EST	20-JUL-1998
LOCUS	D15131/c				
DEFINITION	RICCO129A	Rice callus	Oryza sativa	CDNA clone R,	mRNA sequence.
ACCESSION	D15131				
NID	5286323				
KEYWORDS	D15131.1	GI:286323			
EST.					

SOURCE Oryza sativa.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 328)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP"
Seq primer: oligo(dt).
Location/Qualifiers
1..328
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R"
/note="Vector: Bluescript II SK+; Site.1: SalI; Site.2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dt) as a primer and ligating to the SalI-NotI site of Bluescript II SK+ phagemid.
3 others

BASE COUNT 73 a 100 c 101 g 51 t 3 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

Db 222 TCTCCTTAG 213

RESULT 2
D15980 435 bp mRNA EST 20-JUL-1998
INITIATION RICC1808A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
NID D15980
VERSION 9287175
KEYWORDS D15980.1 GI:287175
SOURCE EST.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 435)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP"

Seq primer: oligo(dt).
Location/Qualifiers
1..435
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R"
/note="Vector: Bluescript II SK+; Site.1: SalI; Site.2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dt) as a primer and ligating to the SalI-NotI site of Bluescript II SK+ phagemid.
1 others

BASE COUNT 108 a 97 c 117 g 112 t 1 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 435;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

Db 70 TCTCCTTAG 79

RESULT 3
D21672 405 bp mRNA EST 07-OCT-1996
LOCUS M818D09 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
DEFINITION clone 81D09, mRNA sequence.
ACCESSION D21672
NID 9618800
VERSION D21672.1 GI:6188800
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 405)
AUTHORS Nishiguchi,S., Joh,T., Horie,K., Zou,Z., Yasunaga,T. and Shimada,K.
TITLE A survey of genes expressed in undifferentiated mouse embryonal carcinoma F9 cells: characterization of low-abundance mRNAs
JOURNAL J. Biochem. 116, 128-139 (1994)
MEDLINE 95096008
COMMENT Contact: Kazunori Shimada
Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.

FEATURES
SOURCE 1..405
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="81D09"
/note="mouse embryonal carcinoma cell line F9"

BASE COUNT 130 a 74 c 117 g 84 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 405;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

Db 157 TCTCCTTAG 148

RESULT 4
D22924 394 bp mRNA EST 20-JUL-1998
LOCUS D22924


```

DEFINITION  RICCI1794A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D22924
NID         9426849
VERSION     D22924.1 GI:426849
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
REFERENCE   1 (bases 1 to 394)
AUTHORS    Sasaki,T. and Minobe,Y.
TITLE      Rice cDNA from callus
JOURNAL     Unpublished (1994)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT = 'RGP'

FEATURES
  source
    1..394
    /organism="Oryza sativa"
    /strain="cultivar Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice callus"
    /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
    NotI; cDNA prepared from rice callus mRNAs by using
    oligo(dT) as a primer and ligating to the SalI-NotI site
    of pBluescript II SK+ phagemid. 4 others
    122 c 117 g 61 t

BASE COUNT  90 a 122 c 117 g 61 t 4 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 394;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    221 TCTCCTTAG 212

RESULT 5
D24072/c 347 bp mRNA EST 20-JUL-1998
LOCUS     RICRI057A Rice root Oryza sativa cDNA clone R, mRNA sequence.
DEFINITION
ACCESSION D24072
NID       9427928
VERSION   D24072.1 GI:427928
KEYWORDS  EST.
SOURCE    Oryza sativa.
ORGANISM  Oryza sativa.
REFERENCE  1 (bases 1 to 347)
AUTHORS   Minobe,Y. and Sasaki,T.
TITLE     Rice cDNA from root
JOURNAL   Unpublished (1995)
COMMENT   Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305

BASE COUNT  129 a 129 c 106 g 113 t 2 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    Db 108 TCTCCTTAG 99

DEFINITION  RICRI932A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D24444
NID         9428296
VERSION     D24444.1 GI:428296
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
REFERENCE   1 (bases 1 to 430)
AUTHORS    Minobe,Y. and Sasaki,T.
TITLE      Rice cDNA from root
JOURNAL     Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT = 'RGP'

FEATURES
  source
    1..430
    /organism="Oryza sativa"
    /strain="Nipponbare, sub_species Japonica"
    /db_xref="taxon:4530"
    /clone="R"
    /clone_1lb="Rice root"
    /note="Prepared from seedling root. "
    129 a 129 c 106 g 113 t 2 others

BASE COUNT  129 a 129 c 106 g 113 t 2 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
    |||||
    Db 108 TCTCCTTAG 99

```



```

RESULT 7
D24993/c 318 bp mRNA EST 20-JUL-1998
LOCUS R1CR2912A Rice root Oryza sativa cDNA clone R, mRNA sequence.
DEFINITION D24993
ACCESSION D24993
NID 9428837
VERSION D24993.1 GI:428637
KEYWORDS EST,
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 318)
Minobe,Y. and Sasaki,T.
Rice cDNA from root
Unpublished (1995)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Yuzo Minobe
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: minobe@rcs0.riken.go.jp
PROJECT "RGP"

FEATURES
SOURCE
location/Qualifiers
1. 318
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone_1ib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 96 a 54 c 81 g 83 t 2 others
ORIGIN

```

Query Match 100.0%; Score 10; DB 20; Length 318;
 Best Local Similarity 100.0%; Pred. No. 4.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 TCTCCTTAG 10
 |||||
 108 TCTCCTTAG 99

RESULT 8
 D25754/c
 LOCUS
 DEFINITION
 accession
 NID
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

D25754
 205 bp mRNA
 HMGS04121 Human colon mucosa Homo sapiens cDNA clone cml88 3',
 mRNA sequence.
 D25754
 g500443
 D25754.1 GI:500443
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 205)
 Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.
 Global analysis of gene expression in colon mucosa: a large scale
 random cDNA sequencing analysis
 Unpublished (1994)

JOURNAL
 COMMENT

Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.
 Institute for Molecular and Cellular Biology
 Osaka University
 3-1 Yamada-oka, Suita, Osaka 565, Japan.

FEATURES	Location/Qualifiers
SOURCE	1. 205 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="cm1888" /clone_id="Human colon mucosa" /note="Adult male, tissue_type = colon mucosa"
BASE COUNT	53 a 44 c 51 g 52 t 5 others
ORIGIN	
Query Match	100.0%; Score 10; DB 20; Length 205;
Best Local Similarity	100.0%; Pred. NO. 4.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
QY 1 TCTCCTTAAG 10 	
DB 118 TCTCCTTAAG 109	
RESULT 9	
D31273	310 bp mRNA EST 08-FEB-1995
LOCUS	D31273
DEFINITION	HUHL13202 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.
ACCESSION	D31273
NID	g644153
VERSION	D31273.1 GI:644153
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 310) Sudo, K., Chihnen, K. and Nakamura, Y. 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library Genomics 24, 276-279 (1995) 95213017
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
COMMENT	

Contact: Iusuke Nakamura
 Institute of Medical Science
 University of Tokyo
 4-6-1, Shirokane-dai, Minato-Ku, Tokyo 108, Japan
 Tel: 81-3-5449-5372
 Fax: 81-3-5449-5433
 Email: yusuke@ims.u-tokyo.ac.jp
 Insert Length: 386 Std Error: 0.00
 High quality sequence stop: 157.
 Location/Qualifiers
 1..310
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal lung"
 67 a 86 c 78 g 78 t 1 others
 BASE COUNT
 ORIGIN
 Query Match
 Best local Similarity 100.0%; Score 10; DB 20; Length 310;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
 QY 1 TCTCCTTAG 10
 |||||||||
 Db 291 TCTCCTTAG 300
 RESULT 10
 LOCUS D34553 414 bp mRNA
 DEFINITION CEAK008B2R Yuj1 Kohara unpublished cDNA Caenorhabditis elegans
 D34553
 ACCESSION D34553
 ID 9522797

VERSION	D34533.1	GI:522797	
KEYWORDS	EST		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae; Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.		
AUTHORS	1 (bases 1 to 414)		
TITLE	Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.		
JOURNAL	Toward an expression map of the C. elegans genome		
COMMENT	Unpublished (1994)		
FEATURES	Source		
Source	Contact: Yuji Kohara Gene Library Lab National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 0559-75-0771 Fax: 0559-75-6240 Email: ykoha@dbj.nig.ac.jp. Location/Qualifiers 1. 414 /organism="Caenorhabditis elegans" /strain="CB1489 him-8(e1489)" /db_xref="taxon:6239" /clone="YK8b2" /clone_1lb="Yuji Kohara unpublished cDNA" /note="dev_stage-varied, sex-Hermaphrodite male, tissue-type=whole animal" BASE COUNT 99 a 92 c 109 g 112 t 2 others ORIGIN		
Query Match	100.0%; Score 10; DB 20; Length 414;		
Best Local Similarity	100.0%; Pred. No. 4.4e+03;		
Matches 10; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 TCTCCTTAG 10		
Db	301 TCTCCTTAG 310		
RESULT 11	D35356	315 bp	MRNA
LOCUS	CELK001DIF Yuji Kohara unpublished cDNA		EST
DEFINITION	Caenorhabditis elegans		08-AUG-1994
SSION	clone yk1d11 5', mRNA sequence.		
VERSION	D35357		
KEYWORDS	EST.		
SOURCE	D35356.1 GI:525357		
ORGANISM	Caenorhabditis elegans.		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae; Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.		
AUTHORS	1 (bases 1 to 315)		
TITLE	Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.		
JOURNAL	Toward an expression map of the C. elegans genome		
COMMENT	Unpublished (1994)		
FEATURES	Source		
Source	Contact: Yuji Kohara Gene Library Lab National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 0559-75-0771 Fax: 0559-75-6240 Email: ykoha@dbj.nig.ac.jp. Location/Qualifiers 1. 315 /organism="Caenorhabditis elegans" /strain="CB1489 him-8(e1489)" /db_xref="taxon:6239" /clone="YK20e5" /clone_1lb="Yuji Kohara unpublished cDNA" /note="dev_stage-varied, sex-Hermaphrodite male, tissue-type=whole animal" BASE COUNT 111 a 77 c 79 g 90 t 3 others ORIGIN		
Query Match	100.0%; Score 10; DB 20; Length 360;		
Best Local Similarity	100.0%; Pred. No. 4.3e+03;		
Matches 10; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 TCTCCTTAG 10		
Db	352 TCTCCTTAG 343		
RESULT 13	D36191	330 bp	MRNA
LOCUS	CELK002GZF Yuji Kohara unpublished cDNA		EST
DEFINITION	Caenorhabditis elegans		08-AUG-1994
ACCESSION	clone yk2g12 5', mRNA sequence.		
NID	D36191		
VERSION	D36191.1		
KEYWORDS	EST.		

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 330)
Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C.elegans genome
Unpublished (1994)

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbi.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..330
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk2g12"
/clone_1lb="Yuji Kohara unpublished cDNA"
/note="dev_stage-varied, sex-Hermaphrodite male,
tissue_type="whole animal"

BASE COUNT
ORIGIN

79 a 89 c 75 g 72 t 15 others

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 330;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

DB 101 TCTCCTTAG 92

RESULT 14
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

D38872 444 bp mRNA EST 11-NOV-1994
R1C1810A Rice callus Oryza sativa cDNA, mRNA sequence.
D38872
9568037
D38872.1 GI:568037
EST.
Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 444)
Sasaki, T., Miyao, A. and Yamamoto, K.
Rice cDNA from callus 1995
Unpublished (1995)

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
Location/Qualifiers

FEATURES
source
1..444
/organism="Oryza sativa"
/strain="Cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_1lb="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:

BASE COUNT
ORIGIN

111 a 96 c 121 g 114 t 2 others

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 444;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

DB 70 TCTCCTTAG 79

RESULT 15
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

D39092 407 bp mRNA EST 11-NOV-1994
R1C1916A Rice root Oryza sativa cDNA, mRNA sequence.
D39092
9568255
D39092.1 GI:568255
EST.
Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 407)
Sasaki, T., Miyao, A. and Yamamoto, K.
Rice cDNA from callus 1995
Unpublished (1995)

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
Location/Qualifiers

FEATURES
source
1..407
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_1lb="Rice root"
/note="Prepared from seedling root."

BASE COUNT
ORIGIN

117 a 70 c 103 g 114 t 3 others

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 407;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCCTTAG 10
|||||

DB 226 TCTCCTTAG 217

Search completed: September 17, 1999, 21:27:55
Job time: 14259 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:00 ; Search time 1962.17 Seconds

(without alignments)
16.208 Million cell updates/sec

Title: US-08-956-518A-87

Perfect score: 10
Sequence: 1 TGTGTCAC 10

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmbl: *
1: gb_dal: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_p11: *
8: gb_p12: *
9: gb_p13: *
10: gb_p14: *
11: gb_p15: *
12: gb_p16: *
13: gb_p17: *
14: gb_p18: *
15: gb_p19: *
16: gb_p20: *
17: gb_p21: *
18: gb_p22: *
19: gb_p23: *
20: gb_p24: *
21: gb_p25: *
22: gb_p26: *
23: gb_p27: *
24: gb_p28: *
25: gb_p29: *
26: gb_p30: *
27: gb_p31: *
28: gb_p32: *
29: gb_p33: *
30: gb_p34: *
31: gb_p35: *
32: gb_p36: *
33: gb_p37: *
34: gb_p38: *
35: gb_p39: *
36: gb_p40: *
37: gb_p41: *
38: gb_p42: *
39: gb_p43: *
40: gb_p44: *
41: gb_p45: *
42: gb_p46: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	5598	1	AAM5961	AJ005961 Acidianus

Result No.	Score	Query Match	Length	DB ID	Description
2	100.0	10498	1	AB008831	AB008831 Acinetobacter
3	100.0	3754	1	AB016431	AB016431 Staphylococcus
4	100.0	10446	1	AE000810	AE000810 Methanobrevibacter
5	100.0	16189	1	AE000845	AE000845 Methanobrevibacter
6	100.0	10072	1	AE000880	AE000880 Methanobrevibacter
7	100.0	30100	1	AF006565	AF006565 Bacillus
8	100.0	29950	1	AF015775	AF015775 Bacillus
9	100.0	2259	1	ASAR60	ASAR60 Anabaena
10	100.0	5198	1	BACWBA	BACWBA Bacillus
11	100.0	2497	1	BACWLB	BACWLB Bacillus
12	100.0	2277	1	BACWYMEI	BACWYMEI Bacillus
13	100.0	8397	1	BSU55043	BSU55043 Bacillus
14	100.0	213190	1	BSUB0004	BSUB0004 Bacillus
15	100.0	208430	1	BSUB0005	BSUB0005 Bacillus
16	100.0	208430	1	BSUB0006	BSUB0006 Bacillus
17	100.0	208430	1	BSUB0007	BSUB0007 Bacillus
18	100.0	208780	1	BSUB0009	BSUB0009 Bacillus
19	100.0	207730	1	BSUB0011	BSUB0011 Bacillus
20	100.0	218410	1	BSUB0015	BSUB0015 Bacillus
21	100.0	212610	1	BSUB0019	BSUB0019 Bacillus
22	100.0	7589	1	BSYLOGCOD	BSYLOGCOD Bacillus
23	100.0	19712	1	D78509	D78509 Bacillus
24	100.0	20341	1	D86418	D86418 Bacillus
25	100.0	17952	1	D90869	D90869 E. coli
26	100.0	1549	1	ECNUPC	ECNUPC E. coli
27	100.0	4756	1	ECNUPC	ECNUPC E. coli
28	100.0	493	1	ECNUPC	ECNUPC E. coli
29	100.0	1997	1	ECNUPC	ECNUPC E. coli
30	100.0	136742	1	ECNUPC	ECNUPC E. coli
31	100.0	2007	1	HPY18984	HPY18984 Helicobacter
32	100.0	2020	1	HPY18987	HPY18987 Helicobacter
33	100.0	1993	1	HPY18988	HPY18988 Helicobacter
34	100.0	1993	1	HPY18989	HPY18989 Helicobacter
35	100.0	2010	1	HPY18990	HPY18990 Helicobacter
36	100.0	6962	1	KPY011907	KPY011907 Klebsiella
37	100.0	4972	1	MAMATRM	MAMATRM Mycobacterium
38	100.0	37821	1	MLCB1770	MLCB1770 Mycobacterium
39	100.0	35445	1	MLCB2052	MLCB2052 Mycobacterium
40	100.0	36310	1	MLCB4	MLCB4 Mycobacterium
41	100.0	3822	1	MLCB4	MLCB4 Mycobacterium
42	100.0	36063	1	MLCB4	MLCB4 Mycobacterium
43	100.0	42106	1	MSGBI133CS	MSGBI133CS Mycobacterium
44	100.0	38721	1	MSGY219	MSGY219 Mycobacterium
45	100.0	209925	42	AF002223	AF002223 Human gen

ALIGNMENTS

RESULT 1
AAM5961
LOCUS
DEFINITION
ACIDIANUS ambivalens schABCD operon, nagd gene, ORF2 and partial ORF1 and 3.
ACCESSION
AJ005961
VERSION
G3378536
NID
AJ005961.1 GI:3378536
KEYWORDS
2Fe-2S-ferredoxin; 4Fe-4S-ferredoxin; flavoprotein; iron-sulfur protein; nagd gene; ORF1; ORF2; ORF3; sdhA gene; schABCD operon; Acidianus ambivalens.
SOURCE
ORGANISM
Acidianus ambivalens
Archaea: Crenarchaeota; Sulfolobales; Acidianus.
REFERENCE
Kletzin, A.
Direct Submission
Submitted (04-MAY-1998) Kletzin A., Institut fuer Mikrobiologie und Genetik, Technische Universitaet Darmstadt, Schnittspahnstrasse 10, Darmstadt, 64297, GERMANY
JOURNAL
2 (bases 1 to 5598)
Kletzin, A.
The succinate dehydrogenase from the extremely thermophilic archaeon Acidianus ambivalens

CEDIKGEHVKGSEVILITDGEKDIAETTVRSLSKDNATLISVMIRGDNALRRVSD
BASE COUNT 1999 a 849 c 1203 g 1547 t
ORIGIN TTVVYKKLDONDLKAYEA"

Query Match 100.0%; Score 10; DB 1; Length 5598;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGTGTGAC 10
|||||
Db 3522 TGTGTGTGAC 3531

RESULT 2
AB008831 10498 bp DNA BCT 05-JUN-1998
LOCUS Acinetobacter sp. genes for aniline dioxygenase subunits and
INITIATION reductase component, partial and complete cds.
ACCESSION AB008831
NID 92627146
VERSION AB008831.1 GI:2627146
KEYWORDS 4-oxalocrotonate decarboxylase; 4-hydroxy-2-oxovalerate aldolase;
acetaldehyde dehydrogenase; 2-oxopent-4-dienate hydratase;
2-hydroxymuconic semialdehyde hydrolase; 2-hydroxymuconic
semialdehyde dehydrogenase; catechol 2,3-dioxygenase; small
ferredoxin-like protein; aniline dioxygenase reductase component;
aniline dioxygenase beta-subunit; aniline dioxygenase
alpha-subunit.

SOURCE Acinetobacter sp. (strain:YAA) DNA.
ORGANISM Acinetobacter sp.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
Moraxellaceae; Acinetobacter.

REFERENCE 1 (bases 1 to 10498)
AUTHORS Takeo,M.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1997) to the DDBJ/EMBL/GenBank databases.
MASSACHUSETTS, Himeji Institute of Technology, Department of
Applied Chemistry, 2167 Shosha, Himeji, Hyogo 671-22, Japan
(E-mail:takeo@chem.eng.himeji-tech.ac.jp, Tel:0792-67-4893,
Fax:0792-67-4891)

REFERENCE 2 (sites)
AUTHORS Takeo,M., Fujii,T., Takenaka,K., and Maeda,Y.
TITLE Cloning and sequencing of a gene cluster for the meta-cleavage
pathway of aniline degradation in Acinetobacter sp. strain YAA
JOURNAL J. Ferment. Bioeng. 85, 514-517 (1998)
3 (sites)
AUTHORS Fujii,T., Takeo,M. and Maeda,Y.
TITLE Plasmid-encoded genes specifying aniline oxidation from
Acinetobacter sp. strain YAA
JOURNAL Microbiology 143 (Pt 1), 93-99 (1997)
9717776

FEATURES
JOURNAL MEDLINE
FEATURES location/Qualifiers
1..10498
/organism="Acinetobacter sp."
/strain="YAA"
/db_xref="taxon:472"
1..276
/gene="atda3"
1..276
/gene="atda4"
/codon_start=1
/transl_table=11
/product="aniline dioxygenase alpha-subunit"
/protein_id="BAA23550.1"
/db_xref="PID:d1024429"
/db_xref="PID:92627147"
/db_xref="GI:2627147"
/translation="ELNATMRTOEDFPIMGEVDVANFESQCEGLTETPEIEMWIDFS
RHMNGENACVQDVVIOHKPTSEIHSRHFEDVILQASVANKENQSEV"
278..874
/gene="atda4"

gene
CDS

gene

CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene

278..874
/gene="atda4"
/codon_start=1
/transl_table=11
/product="aniline dioxygenase beta-subunit"
/protein_id="BAA23551.1"
/db_xref="PID:d1024430"
/db_xref="PID:92627148"
/db_xref="GI:2627148"
/translation="MNNKDLISIKIVDESAVNLAVYOIEIKOYSDYFNLSNIGEPALD
HKINMFLKREARLDQCEDEMTLFTLEDGCGYWIPIGSPAPASPASEAATPEHDIRRK
DRIYRQGFAPVSOQPVSKTRILGAPRYMAVPPSSSEGLVTRTSITVYESDGRSQVL
SGWIGYVILKQNDLKNMOKIINLNDCLSPGNNSEFL"
885..1895
/gene="atda5"
885..1895
/gene="atda5"
/codon_start=1
/transl_table=11
/product="aniline dioxygenase reductase component"
/protein_id="BAA23552.1"
/db_xref="PID:d1024431"
/db_xref="PID:92627149"
/db_xref="GI:2627149"
/translation="MNTLKRVIDKIAETKESFVLKPLDGVLAHSPGYLPIKIR
TEKGLFRSYSLSSASANEDFKITVKEKRGGRSNMCDNVKAGDFIETLPPASFSH
PQNDROFVFAFGSGGIPVSIKITALNRKRIKFLVANSSESSIFRKLKDLCL
QEPRLDLOFLDEKGIPTSIAREQYIDDALEVEYFGAPFPMGVEENLIESKVP
PLITKEFASVSDNDNGDTESSAEADQVYVFMNLNGIKNSVMSCEDPFIETIKAG
INPSSCCAGNCSCMLVBSGDVILBSNTVLDASDEDDGILACRSFRSKNIETISF
DQ"
1902..2828
/gene="atda6"
1902..2828
/gene="atda6"
/note="similar to LysR-type transcriptional activator"
/codon_start=1
/transl_table=11
/product="small ferredoxin-like protein"
/protein_id="BAA23553.1"
/db_xref="PID:d1024432"
/db_xref="PID:92627150"
/db_xref="GI:2627150"
/translation="MRCVLKHSSTYDNAMNDVKAFLYRAKDYEAASESLGVDGST
LRRKIQSLLEVVLGSLFYKNNRWALPGLDEVKAAVDMESATRLFGILPGRGVY
IKVSIPELVNQFSDVLAIFKIPYEFTEFSDARFVDELEDEGFAIRLPGRGVY
NSLKIRGTGGLGVFSGASVYIDIVDFEGDALRDSDLTKTGIGSYKRDVPGIL
LDWQDLGFGYNIKIMCDLDESCARFCBAGLALILPKFLAHRYFMNSCVIDASDTLMA
ELNLSRDKMSDMQVVLGNLAKSKEMEVS"
2888..3226
/gene="atda7"
2888..3226
/gene="atda7"
/codon_start=1
/transl_table=11
/product="small ferredoxin-like protein"
/protein_id="BAA23554.1"
/db_xref="PID:d1024433"
/db_xref="PID:92627151"
/db_xref="GI:2627151"
/translation="MGESYQITTEOCSQORPPCKAGOSVLKAMEDOGLGECAPVGRGGG
CGLCKVTVREGDYCEGKMSRVHAPPALAOGEVLACRIYPLSLDILIECROPSAAGLA
NEYTTKAMR"
3232..4155
/gene="atda8"
3232..4155
/gene="atda8"
/codon_start=1
/transl_table=11
/product="catechol 2,3-dioxygenase"
/protein_id="BAA23555.1"
/db_xref="PID:d1024434"
/db_xref="PID:92627152"
/db_xref="GI:2627152"


```

Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 TGTGTCACAG 10
          |||||||||
Db 10167 TGTGTCACAG 10176

RESULT   3
AB016431
LOCUS    AB016431       3754 bp        DNA            BCT           06-FEB-1999
DEFINITION
Staphylococcus aureus, zinc responsible operon czt genes, complete
and partial cds.
ACCESSION
AB016431
NID      94126670
VERSION  AB016431.1 GI:4126670
KEYWORDS
cztB; cztA.
SOURCE   Staphylococcus aureus (strain:912) DNA.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Staphylococcus.
REFERENCE
1 (bases 1 to 3754)
AUTHORS Ohta,T., Kuroda,M. and Hayashi,H.
TITLE    Direct Submission
JOURNAL  Submitted (21-JUN-1998) to the DDBJ/EMBL/Genbank databases. Toshiko
Ohta, University of Tsukuba, College of Medical Technology;
Tenodai 1-1-1, Tsukuba, Ibaraki 305, Japan
(E-mail:tohtaisakura.cc.tsukuba.ac.jp, Tel:+81-298-53-3454,
Fax:+81-298-53-3454)
2 (sites)
AUTHORS Kuroda,M., Hayashi,H. and Ohta,T.
TITLE    Chromosome-determined zinc responsible operon czt in Staphylococcus
aureus strain 912
JOURNAL  Microbiol. Immunol. (1999) In press
FEATURES
Location/Qualifiers
source     1..3754
            /organism="Staphylococcus aureus"
            /strain="912"
            /db_xref="taxon:1280"
            /complement(<1..458)
            /note="Hypothetical protein"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA36685.1"
            /db_xref="PID:d1037672"
            /db_xref="PID:g4126671"
            /db_xref="GI:4126671"
            /translation="MFTYSAKFNKKPOLLTITLYALIAFAVIAVIYLANFOLAKRY
AQRTAIFKRGFGOKPVDAYLKVAVYLIAVVSFLVVOIFGITNVWKRAMEXKYK
FTDFEIKFRKGSLVKLVGLVSIAMIVLSLTSLKLKSPSEMINMS"
            complement(466..471)
            complement(671..716)
            complement(724..1701)
            /gene="cztB"
            /complement(724..1701)
            /gene="cztR"
            /function="Zinc resistance"
            /note="czcd"
            /codon_start=1
            /transl_table=11
            /protein_id="BAA36686.1"
            /db_xref="PID:d1037673"
            /db_xref="PID:g4126672"
            /db_xref="GI:4126672"
            /translation="MSSHSHDHMHSHVTDNKKYLFISFILGLVFIEIIGGLAN
SLLNSGIMESDPTSGVALVAFTYEKNATTKTGFGRFEVLNALNGTLEVI
SILTVBARIKRFVEPEVOSKEMLIISIIGLVINIVAEFMEFKGGDTSHNLNRGAFL
HVIGDIIVSGVATAIAILMAWGWTIADIPASIVSAIVSIIKSAMGITSNNILMGST
PSVSDIDEGVITTKKDSRIOSVHDCHVYTISDNKAISCHEVVYDHTLTMCCELLLEN
IEEDLLHLNIHNHTIQLTEPPNKHDEHSITICGSTHSHNHHAHHAHVH"
            complement(1703..2023)
            /gene="CztA"
            /complement(1703..2023)

```



```

/gene="CzrA"
/function="repressor"
/codon_start=1
/transl_table=11
/protein_id="BA36687.1"
/db_xref="PID:d1037674"
/db_xref="PID:g4126673"
/db_xref="GI:4126673"
/translation="MSEQSEINDTLEKTEIFKALGDYRIRIMELLSVSAVGH
ISHQINLSQSNVSHQKLKLSVHLVKARQSQSMIYSLDDIHVATMLKQAIHNNHFK
ESGL"
complement(1709..1714)
/gene="CzrA"
complement(2034..2039)
complement(2064..2069)
complement(2087..2092)
2161..2166
2175..2813
/feature="Hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BA36688.1"
/db_xref="PID:d1037675"
/db_xref="PID:g4126674"
/db_xref="GI:4126674"
/translation="MNLIVIGANGVGSLLVQOLAKENVPPTAGVROSQDQINLKSOG
AKALLVENDSIETLETETKPFDPKVFVSQSGNTGADKTIIVLDGAVKSMASKE
ANIKHYVSTYDSRQAFDSDGLKPYTIKAKHYADYLRSGSLIHHGALTNAA
GSGRIEAAQYDFDGEIREDYATVLEKIYVSNHNFQEIISGEDIDKA"
3263..3268
3275..>3754
/feature="Hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BA36689.1"
/db_xref="PID:d1037676"
/db_xref="PID:g4126675"
/db_xref="GI:4126675"
/translation="MPLFLQILKTKLWGQRLSEFGYOLDNDITGECWCVASHPNGT
SEINGPYQGLDRIWSEHRELEGEDEPSKDFLTKIIVDARESLIHHPDNSYAE
HENGQYKSECWYIIDAEDAEIYIGTIAESREYANVHQGTIESILRYIKVAPGEF"
BASE COUNT 1348 a 534 c 666 g 1206 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 3754;
Jest Local Similarity 100.0%; Pred. NO. 2.7e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGTGTGTGAG 10
|||||
DB 3380 TGTGTGTGAG 3389

```

RESULT 4
AE000810 10446 bp DNA BCT 16-Apr-1998
LOCUS Methanobacterium thermoautotrophicum from bases 172512 to 182957
DEFINITION (section 16 of 148) of the complete genome.
ACCESSION AE000810 AE000668
NID 92621277
VERSION AE000810.1 GI:2621277
KEYWORDS
SOURCE
ORANISM Methanobacterium thermoautotrophicum.
Methanobacterium thermoautotrophicum.
Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
Methanobacterium.
1 (bases 1 to 10446)
Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B.,
Qiu,D., Spadofora,R., Vicare,R., Wang,Y., Wierzbowski,J.,

```

TITLE
JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
Gibson,R., Jivani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,
Prabhakar,S., McDougall,S., Shiner,G., Goyal,A., Pietrowski,S.,
Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nolling,J. and
Reve,J.N.
Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics
J. Bacteriol. 179 (22), 7135-7155 (1997)
9603/514
2 (bases 1 to 10446)
Smith,D.R.
Direct Submission
Submitted (10-ANG-1997) Genomics and Technology Development, Genome
Pharmaceutics Corporation, 100 Beaver Street, Waltham, MA
02154-8448, USA
1..10446
Location/Qualifiers
/organism="Methanobacterium thermoautotrophicum"
/strain="delta H"
/db_xref="taxon:2166"
/clone="MTH"
/complement(86..796)
/gene="MTH230"
/complement(86..796)
/gene="MTH230"
/feature="Function Code:14.00 - Unknown. : similar to,
gp:GI:91652373 LN:D50905, p(-)=0.999, pid=11%"
/translation="unknown"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA84736.1"
/db_xref="PID:g2621278"
/db_xref="GI:2621278"
/translation="MDTKKIIIGLFIIVILGSVYENYITNSONFVGEKHYLLIC
ADGEPGIGAVDMAEITIMNEGNTINVAVPYPMNAHPTAQSPLRAQVSRML
HDLWENDTERGARLAEIYENYNTGKIDIVVTPOAVDGILOSIPVYEGQYVS
GNSIOPLREQOQSSRGVALQSLMRAIFMATDRSKYLMVNAAGLQYQGNIVYP
EGAVOPLISTGERVGN"
1053..1781
/gene="MTH231"
1053..1781
/gene="MTH231"
/feature="Function Code:14.01 - Unknown. Conserved protein;
similar to, pIR:LN:B64471 AC:B64471, p(-)=3.1E-32, pid=37%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AA84737.1"
/db_xref="PID:g2621279"
/db_xref="GI:2621279"
/translation="MYSSEGAMRIYAGVGENNMERAA5LAEVDLVHSEEFIEE
LRGAAYVRSLSPLANIMAEIKKGLPLNLSAVIEVGANFLAPVIGDGRVDDRF
KIAVASSEFLRTGTEPRVGVISGGRGDRSPEDRSIHEEPLTSMKDKYRVRH
YHLLIEAVADGCVNIIAPDITGNLIFRSVLVGLARSGAVALGDFIVDTSSQ
TAGYLRLAKLFAHMLARGNEDNE"
1858..2625
/gene="MTH232"
1858..2625
/gene="MTH232"
/feature="Function Code:14.01 - Unknown. Conserved protein;
similar to, sp:LN:ID72_METUA AC:058767, p(-)=1.6E-72,
pid=53%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AA84738.1"
/db_xref="PID:g2621280"
/db_xref="GI:2621280"
/translation="NSPLPLKYLYEYISRLRRDRMPHVAIINDGNRRYKLOGS
MNPIDGKGIETLEKVDLDCVLEIYTAFAFSTENRFRKPEKVKLKEPNEF
AIA5NRKIKNRVRAVAGKLELDPDVRALITAEKSTROYSDRLVNAIGDGOE
IYATKRIADYVAAGLIDPEDIDEDINRRLITAGLEDPLILITRTSGEHLGSFLMQ
SSISELYDQSLVPELNRKVDPLRAINSYQORRRFV"
2682..3452

```

gene

gene
CDS
/gene="MTH233"
2682.3452
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pIR:LN:E64497 AC:E64497, p(-)8.6E-44, pld=36"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB84739.1"
/db_xref="pid:g2621281"
/db_xref="gi:2621281"
/translation="MGGIITIVHCHLDFKDFNRNKEEIERARSKLRAVIDSGVGLG
GRRRLLESLNPGICPTMGHPVDASAKADLIGVVSQIESNTDLIAVGETGM
FHRTDEGRROEETFRVFEELAEHEMPLVVARDAERALEVLEVRPEVIFHC
YGSIEIARLIDEGYIISITLVAFSEHHEMLVRAIPLEGMLTEDSYLSPPFGKR
NEPAVEAVARAIARIKMDLEDVDSITANMERVGL"
complement(3518.3895)
/gene="MTH234"
complement(3518.3895)
/note="Function Code:13.07 - Other, Unclassified; similar
to, sp:LN:DC4C_ACICA AC:P20370, p(-)1.2E-12, pld=288"
/codon_start=1
/transl_table=11
/product="gamma-carboxymuconolactone decarboxylase"
/protein_id="AAB84740.1"
/db_xref="pid:g2621283"
/db_xref="gi:2621282"
/translation="MERYRKGMEILNRMRKSYTAIRDELVDVADPLARFVAPVAGD
YSRVGLDKIRLELTLALVLRADDDQLSKHVGALNKGCSKDEITEVIMQAVYAG
EPAAINAVLAAAEVETENDPAEV"
3989.4639
/gene="MTH235"
3989.4639
/note="Function Code:9.02 - Metabolism of Cofactors and
vitamins, Riboflavin metabolism (B2); similar to,
pIR:LN:G64383 AC:G64383, p(-)3.8E-50, pld=498"
/codon_start=1
/transl_table=11
/product="riboflavin-specific deaminase"
/protein_id="AAB84741.1"
/db_xref="pid:g2621283"
/db_xref="gi:2621283"
/translation="MRPYILNAAATLDGKIATATGSGESGEEEDLRVHELRRECD
IMVGINVLAADPRLTVHRVDAAPGDNPRVYVDSMARPPFRVILNDEAPVIGVSE
SAPPERVAELRRAEVVAVATRRVDLHLLERLHMGIRLMBEGSTLINSMLTGL
VDEVYCIAPLVIGGRDARTLVDSGIDEMADAIRLELKRSTLTGLEDLIVETVKG"
4642.5187
/gene="MTH236"
4642.5187
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pIR:LN:A64438 AC:A64438, p(-)8.3E-46, pld=488"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB84742.1"
/db_xref="pid:g2621284"
/db_xref="gi:2621284"
/translation="VFMLEDIYERIVIRIEGRCRECKVVCRRMDFOFNQMSRLDL
OETISRSPPVRPALDPMISTGVRDNDENRLLGYPCCIRSFSENPVRAIDG
EHLAEVSELDIEGKCAIMPSCFIPSCRCCEAMERKLIIGADDERRRIELEDEL
NMRLPHFLVAYEVEFKYLE"
5346.9311
/gene="MTH237"
5346.9311
/note="Function Code:9.10 - Metabolism of Cofactors and
vitamins, Porphyrin and chlorophyll metabolism; similar
to, pIR:LN:S64721 AC:S64721, p(-)1E-97, pld=248"
/codon_start=1
/transl_table=11

gene
CDS
/product="magnesium cheletase subunit"
/protein_id="AAB84743.1"
/db_xref="pid:g2621285"
/db_xref="gi:2621285"
/translation="WNDRSOGGENMKRFLIALILLISVPSVSAADNNTTSNTES
KMLKDTENVLLTGVGVTPDPMNEAVRNDLPAVGFYTLKTYMDFNINSASTQ
FKNDKTVDFIEFTFRPGYPTGMSGEPALVDLOIAASLRKGAIFVIGPKPN
ITGVVNTLPAVGPVAPALSRNKRKIVLHSGVNTLANTDKIPVMOEPIV
PLAPOVYDRAPYMDMNTLIRPGARVAVAILNRYLSGNDYETTLQKLESRG
LNVITFYCSDPIGASRFEVWNTSVDAVACVQFQYWDNQTITFTDLNPAVOG
PLPEFLQSLDDYINSTROQLGLEVYMLFEMEQLEIPILIGSRIQSDPLTGV
TIKEVAYEPGIDQVNRITLAWNLRRKANSDFKLAAMYEPSTHEDGAPANGNIN
SLSNLLMRAPAGSTGNGNTEIYELLIRASGNPNMOTELARKVANGCITIPV
SEYLRKYGASLHSGVYEAALMGPPGIMINNSVYPIGMGMLILAPOVKNWG
TLNNTLPTTHQIFAPYXLOKGCADAVYICQGTIELLPGHNNMTEEDWNTLI
GSRPHTLKNEPDEAVINFTKRAIVITISYLPVWRLELGVQELSDLMGYSIS
TAESGDTDRNNSLESLTRDVTFRAGLEVRLGVNSTAFSVLRLRHEYLELSGILT
PYGLTFEGLPDSLETFIDALISFDPAANTARDEIRNLIMSASNNLSLALN
GEFIQVARSVPVINTLELPTGRNNYTPDPSIPDPAVVGSRRAEMLKRRNSG
GRPEYVAVDIDGVSTGGOSIAIEFYLRPIYESGALIGTEIIPLELGRPRIDY
VISDPHNRGAIPATMDYINDTIRIVNLNSSEMYRKRKYLANKAGIYELVAGM
NTSEADANADRLARTIRIGLPADPHAGVDRILMSKSDWTPQOLATYLSYIAY
GRDLNGVQSPKLESLELRTVDTVWISPYRAGGCTCLRVSVTVNEVNTLGRNIN
SYIARTAGTPLIRLOSITYDVLAVILNVAWQKIREGSGSALQVRLTS
DALVDVAPDVWRIADTFLPDSVSROSDSALOMIARYROAHTRLGLVLSQELV
AISEMLGERSGSDGDSGTTTPHGTTGSGATGGGRTGCTSPGLSPVSGAISMDSO
SSGASASPGSGERAGRSYISIPTAKSSGTQLXALNIGITGLPCLLGVGYTGPMR
K"
complement(9445.10365)
/gene="MTH238"
complement(9445.10365)
/note="Function Code:2.02 - Energy Metabolism, Methane
metabolism; similar to, pIR:LN:B64382 AC:B64382,
p(-)7.3E-70, pld=318"
/codon_start=1
/transl_table=11
/product="lungsten formylmethanofuran dehydrogenase,
Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 10446;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 840 TGTGTCTCAG 10
1 TGTGTCTCAG 10
|||||||
Db 840 TGTGTCTCAG 849
RESULT 5
AE000845 16189 bp DNA BCT 16-APR-1998
LOCUS Methanobacterium thermoautotrophicum from bases 571179 to 587367
(section 51 of 148) of the complete genome.
ACCESSION AE000845 AE000666
NID 92621726
VERSION AE000845.1 GI:2621726
SOURCE Methanobacterium thermoautotrophicum.
ORGANISM Methanobacterium thermoautotrophicum
Archaea: Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
Methanobacterium.
REFERENCE 1 (bases 1 to 16189)
Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M.,
Dubois J., Aldredge T., Bashirzadeh R., Blakey D., Cook R.,
Gilbert K., Harrison D., Hoang L., Keagle P., Lum W., Pothier B.,
Qiu D., Spadafora R., Vicare R., Wang Y., Wierzbowski J.,
Gidson R., Jiwani N., Caruso A., Bush D., Saefer H., Patwell D.,
Prabhakar S., McDougall S., Shimer G., Goyal A., Pletrowski S.,
Church G.M., Daniels C.J., Mao J.-i., Rice P., Nolling J. and
Reeve J.N.
TITLE Complete genome sequence of Methanobacterium thermoautotrophicum
strain: functional analysis and comparative genomics
JOURNAL J. Bacteriol. 179 (22), 7135-7155 (1997)

MDLINE 98037514
REFERENCE 2 (bases 1 to 16189)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1997) Genomics and Technology Development, Genome
 Therapeutics Corporation, 100 Beaver Street, Waltham, MA
 02154-8448, USA

FEATURES
 source
 Location/Qualifiers
 1..16189
 /organism="Methanobacterium thermoautotrophicum"
 /strain="delta H"
 /db_xref="taxon:2166"
 /clone="MTH"
 complement(103..1338)
 /gene="MTH645"
 complement(103..1338)
 /gene="MTH645"
 /note="Function Code:10.12 - Metabolism of Macromolecules,
 degradation of proteins--peptides--and glycopeptides ;
 similar to, sp:LN:1090_MetUA AC:057555, p()=1.3E-88,
 pid=46%"
 /codon_start=1
 /transl_table=11
 /product="collagenase"
 /protein_id="AAB85150.1"
 /db_xref="PID:g2621727"
 /db_xref="GI:2621727"
 /translation="MLYKIKDCFSVIEPMLSPAGDFSLMALRSGADAVYIGLEG
 CNRARADNFRVSEVAEVAERADHGSRLVCNTALRDDVRLKNIPELHSGAVD
 AVASDLAVALAEGGLDVHVSQANITRTLTILRDMGVSRALISRELISREIAG
 IASAPVEVDFPHGALCTRAISRCYLSLTYRSANCGDCLPCPKSRNLSDEDF
 DLEVSASVAVSTYLSPADLCMEH1PELDAGVHALKIEGRGRADYATVGVRE
 ALDRYISGEWREFEERWLSERKVFENRFGFTGFSEPGNSGNISEYIKEDIGEVN
 YVRKVGAAARLRLMRPLRVGDEVIIQCKTGALLIQVKSQIEGESVEVAEGGVGLV
 EDVRPGDLYRRVKKRPPDS"
 complement(1363..2787)
 /gene="MTH646"
 complement(1363..2787)
 /gene="MTH646"
 /note="Function Code:4.01 - Nucleotide Metabolism, Purine
 metabolism ; similar to, sp:LN:PURL_MetUA AC:057557,
 p()=6.2E-121, pid=52%"
 /codon_start=1
 /transl_table=11
 /product="amidophosphoribosyltransferase"
 /protein_id="AAB85151.1"
 /db_xref="PID:g2621728"
 /db_xref="GI:2621728"
 /translation="MLGESEVRDKGIVIGYSQDKKTGVAQIYVALYALQHRQESA
 GISTFNGNDILTHRGGLVCDVNPKELELKGAVGIGHVRYSTGESIENQPFWS
 EFOGKIAIAHNDIINSMLELEEGHNFSTDEYICHLISREDEKRMYS
 IKVSEOLVGSYLVLLNODLVYVDPVGIKPLAFARSGSTOIVASEVAPVIGAE
 HVNDVPGETLHNRRKSYVANAQNTNRARHCFEYVYRARPVSVDGRNRYVRNTI
 GELVIREHPANADVVPVDPDSIPALIGISRASGIPYGBLNNRYGRTFLPTCEE
 RETAVKLNKMPNIRSELEGRKRIYIDISIVNGTIRALDIIRDAGAEHLRGCPPI
 KSPCYGIAMATKEELIASTRANVEEIRIIGVDSLGLYSIESIVECIGIKKGLCTGC
 LDDSDYPTLPSPDISEXEMRSCSR"
 complement(2803..3645)
 /gene="MTH647"
 complement(2803..3645)
 /gene="MTH647"
 /note="Function Code:14.00 - Unknown, ; similar to,
 pIR:LN:s61081 AC:s61081:s60817, p()=0.0098, pid=10%"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAB85152.1"
 /db_xref="PID:g2621729"
 /db_xref="GI:2621729"
 /translation="MDNKGPDVDRITVEGASDVESYRALQRYSGAKYHTISSIVP
 TTSLEIALRAVEGADIVLTVDVDTQGRRELDLFRRLRGHVGHITRMLPYGHDEV
 LDDPLIRELEIENAIIRAGLDTLIGISLSMKRLECRKRLDEYVAETALDENTR
 LQGEVAGNRIEISLNGELSQLBEKFRLLLEGESKLETRSELEDEKELLTETSITDLW

gene
CDS
 RETPEEPDLEIRIFVTDHKEGIIIGGSIAPSRDAVENLRINSAVLTEND
 EESS"
 complement(3646..3828)
 /gene="MTH648"
 complement(3646..3828)
 /gene="MTH648"
 /note="Function Code:10.04 - Metabolism of Macromolecules,
 Ribosomal proteins; similar to, sp:LN:RL37_MetUA
 AC:P54011, p()=2.9E-20, pid=68%"
 /codon_start=1
 /transl_table=11
 /product="ribosomal protein L37"
 /protein_id="AAB85153.1"
 /db_xref="PID:g2621730"
 /db_xref="GI:2621730"
 /translation="MKGTSPFGKRNKNLHRCRCGKNSYHVKKVCAACGFRSHRI
 RYSWONKRTGGRK"
 complement(3894..4139)
 /gene="MTH649"
 complement(3894..4139)
 /gene="MTH649"
 /note="Function Code:14.01 - Unknown, Conserved protein;
 similar to, gp:GI:g495021, p()=1.5E-09, pid=35%"
 /codon_start=1
 /transl_table=11
 /product="conserved protein"
 /protein_id="AAB85154.1"
 /db_xref="PID:g2621731"
 /db_xref="GI:2621731"
 /translation="MIDVSSORVNOVRPDLGNSLNSPVIIKKNDRFRGVLKSF
 LHMNLVNDAAELEDGEVTRRLGYLIRKDNIVYISP"
 4441..4923
 /gene="MTH650"
 4441..4923
 /gene="MTH650"
 /note="Function Code:14.01 - Unknown, Conserved protein;
 similar to, pIR:LN:G64478 AC:G64478, p()=3.3E-35, pid=44%"
 /codon_start=1
 /transl_table=11
 /product="conserved protein"
 /protein_id="AAB85155.1"
 /db_xref="PID:g2621732"
 /db_xref="GI:2621732"
 /translation="MTVKIRKRYHIIKKKRLREIRGOIGDYSSLPEKATVEIETD
 NIIIVNGEPLNFIIDRVPYTLGALKMIDIGYVVDGAVFLNAGDVAVSPGIYD
 APEISENTVYVVDERNRPLAVGISLSGPEMERDGSKAVKTIHHIGDAIMELEV
 "

gene
CDS
 4923..5627
 /gene="MTH651"
 4923..5627
 /gene="MTH651"
 /note="Function Code:14.01 - Unknown, Conserved protein;
 similar to, sp:LN:Y116_MetUA AC:057580, p()=4.6E-45,
 pid=40%"
 /codon_start=1
 /transl_table=11
 /product="conserved protein"
 /protein_id="AAB85156.1"
 /db_xref="PID:g2621733"
 /db_xref="GI:2621733"
 /translation="MLVELINDAGNITISDVHIGITLAVGSHLENHGPALPTDQAKI
 ASYVALALRTGARELGLVLAASEPYVKGIIHMDRDLVRELPVARKARLNL
 EAAVYVNGKNGKDCMDLEEGLEIETAMNRYVEIGSPHAGSELSAGLITIGAD
 IRRDCEIPELYPEIGMIGLKKARRENREIDAAKICENDGVNPDVYLGRLIDDAVE
 SYISDKELKIIQEP"
 5635..7746
 /gene="MTH652"
 5635..7746
 /gene="MTH652"
 /note="Function Code:4.01 - Nucleotide Metabolism, Purine
 metabolism ; similar to, gp:GI:1657993 LN:TAU73619,
 p()=5.7E-145, pid=44%"
 /codon_start=1

/transl_table=11
/product="ribonucleotide reductase, large subunit"
/protein_id="AAB85157.1"
/db_xref="PID:g2621734"
/db_xref="GI:2621734"
/translation="MISLSTALVKLBERYLLRGEQGEVETPEMFRVRAVAVASD
EAYDDPAAVEAATYVWNGIEFPNSPTLNNACTPINOASACVLPVEDSIDSES
LRDAIILHKSIGSGVFSRLPRGDIVASTVAVSGVSPFMRLEVAADVIKQGR
RGAMGVLVHSHPDIFSEIDAKSREGPLRNENLSVAPDEMDRVPVDLNPNGEV
DSVSRVILKRIIVAAWSDPGILFEDIRINRNPPTOLGRIEATNPGVSDITVMS
GGPRVLELEKPTALIRGSGPCPGGFETCRDYDLRTREGHCLRLHHRVIV
MDGLFWRAGELRGDRIVMDAAGEPALATFRGLRGAGRODVADATVYGAAPFA
NGFIYHNGEPTLTHESCNISGYNLSMSPSSINNEKLRITHVAVHEDNYDA
SYLPVDEMTLRTKIGLVGFADMLIKGITVNSVALEVAGVMSFTSSSARA
SMELARSGSPPEKGSIMDIQGECKRNATLTITAPTGLSIIAGTSSGILEAFV
FTRNILGRSEHLEPFTAGRLDKRSLERISGLRGVGPAPARIRLEVTAAHEI
DPVHVMAQAFQRYDVNAVKTVNLPADESPADVENVRAHGLGCKGVTVYKSGK
VDEYLRPEVAGSCDMTCBY"
gene
complement(8068. .9591)
/gene="MTH653"
/complement(8068. .9591)
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, gp:GI:g1272330 LN:MTU51624, P()=3.9E-46,
pid=248"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85158.1"
/db_xref="PID:g2621735"
/db_xref="GI:2621735"
Query Match 100.0%; Score 10; DB 1; Length 16189;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGTGTCTCAG 10
Db 2616 TGTGTCTCAG 2625

RESULT 6
AE000880/c
LOCUS
DEFINITION Methanobacterium thermoautotrophicum from bases 1008430 to 1018501
(section 86 of 148) of the complete genome.
ACCESSION
AE000880 AE000666
g2622192
ION
AE000880.1 GI:2622192
SOURCE
Methanobacterium thermoautotrophicum.
Methanobacterium thermoautotrophicum
ORGANISM
Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
Methanobacterium.
REFERENCE
1 (bases 1 to 10072)
Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothler,B.,
Olu,D., Spadafora,R., Vicare,R., Wang,Y., Wierzbowski,J.,
Gibson,R., Jivani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,
Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pletrovski,S.,
Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nolling,J. and
Reeve,J.N.
Complete genome sequence of Methanobacterium thermoautotrophicum
delta: functional analysis and comparative genomics
J. Bacteriol. 179 (22), 7135-7155 (1997)
98037514
2 (bases 1 to 10072)
Smith,D.R.
REFERENCE
JOURNAL
MEDLINE
TITLE
AUTHORS
JOURNAL
Submitted (10-AUG-1997) Genomics and Technology Development, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA
02154-8448, USA

FEATURES
source
Location/Qualifiers
1. 10072
/organism="Methanobacterium thermoautotrophicum"
/strain="delta H"
/db_xref="taxon:2166"
/clone="MTH"
/complement(277. .1050)
/gene="MTH1092"
complement(277. .1050)
/gene="MTH1092"
/note="Function Code:12.11 - Cell Processes,
Decarboxylation; similar to, sp:LN:YC38-CTAPA AC:P48278,
P()=6.5E-23, pid=268"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAB85581.1"
/db_xref="PID:g2622193"
/db_xref="GI:2622193"
/translation="MIMAEIEGIVTLMREKRPFRSRIYTSIVPLMLIFGT
GLGAARFGVPGGYKATYPGIIQGVTFSEGSVYIDROYGLKEILVAPIS
ESMVGKALGISPMIOAILLALSFIVGIMSPOCFQWAVIALIISMGIGGIV
IAFTVMEGFNINSEFTYLPITLSGALPPTIGDPAWQGAVTINPLTYADALRFT
ILRSVLPLEVNILVTITFAVILVLAFLFNKEQNL"
gene
complement(1047. .1985)
/gene="MTH1093"
complement(1047. .1985)
/note="dauonocubcin resistance); Function Code:12.02 -
Cell Processes, Transport of carbohydrates organic acids
alcohols and lipids; similar to, gp:GI:e283950:91707742,
P()=3.3E-67, pid=318"
/codon_start=1
/transl_table=11
/product="abc transporter (ATP-binding)"
/protein_id="AAB85582.1"
/db_xref="PID:g2622194"
/db_xref="GI:2622194"
/translation="MDHIEIEDIKKYEDEFIAVKGVNLRVEGSIYGLVGNAGKT
TILSMCLTILRPTGRTGVNGYDVARAKVRESIGVOSRALDILHGRLEKHA
ATGVPRDVRDRRIEVELIALGDKADSVTVTSGGMRIELIGGLIHHRVLEFD
EPTLGDPTRESITARYTEKLNREEDVTYLLTHHYEEDKDKLCEPAVTSRGETIKAD
SPENLKEGADITIVVDARGFHEILKQDVKAYALMDDEVKLVVERGENLVPEI
VNLAANNVYRVAVLEHPTLEDVFKYGRGISEA"
gene
2082. .3494
/gene="MTH1094"
2082. .3494
/gene="MTH1094"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, gp:GI:g1653005 LN:D90910, P()=1.4E-98,
pid=458"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85583.1"
/db_xref="PID:g2622195"
/db_xref="GI:2622195"
/translation="MDNFETKAPKNGIKGLDMITEGPPRGNGTILYGGPGCKGPIA
MEPLLGAAYVEPGVYSPFDMENMLINPSSDLRLDEGLLEPDSRGKQPIA
DNASISLEKLELEDAIRTGARVYLDKVNLEPGRQCHIGLELLELRWLMG
GVATSVTSGDYGGSPHRLLEYISDCVHLHTPEQVGTRLRIYKVGSGHGLNRY
PFIITRGASIEPTISIDSVSRVSTGIPIDEMGVGTRLRIYKVGSGHGLNRY
KTLISKFAVESCRRGECILFNSNEPAQIVRNMSISIKLGEIFAGDLLHSRPT
SIGLEHLVYMODLMDPNDSVIVPVGLAGAGGSPETRNENAKHLEFIRLDFLKG
KGTSTSPFSLISPFATATTELEKSLSDITWYVLESIRANGERTSRLIRKRGANHS
SSVAEYRFTDRGILLKGGSW"
gene
3504. .3758
/gene="MTH1095"
3504. .3758
/gene="MTH1095"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, gp:GI:g1653851 LN:D90917, P()=0.000000003,
pid=358"


```

gene
CDS
    /translation="MWAGMPSOPDQKSPKSPYILNFIINPMTGSDILGLMAYVYA
    VILLVEERLQGRPNLSKRVHVMNGNLIPIELFSRVALYITLAAPIILITLSP
    YSLRANKHRASSTGHGIGLVYISITPTIATLTFPEAWITGIGIAASVGDPSRLTG
    ERGRTEFTSVLQKSLSESLGMFILLVLPVLTLYTSOGFTPYILLIGVALVSTYLE
    ALPRLQIDNLTAFCGAAVAYILGGMGI"
    /gene=".7461
    6871..7461
    /gene="MTH1100"
    /note="Function Code:14.01 - Unknown, Conserved protein;
    similar to, sp:LN:KTHY_METUA AC:057741, P()=0.0000032,
    pid=19%"
    /codon_start=1
    /transl_table=11
    /product="conserved protein"
    /protein_id="AB85589.1"
    /db_xref="PID:g2622201"
    /db_xref="GI:2622201"
    /translation="MIDGLDGAKNDTHAEILRRRYLERGHVYFRSPEDDNPYGRKA
    REMLDGKGVNHIRAIFALVDIRSLMTLHWSNRPDILFSRYLMGVAAYDPGLSS
    SVYRLRSRLPTPEYFMDVSPESLRLMRMRDEHEMENDDLSTREKALRLAD
    GMYIINTEPIEYQSRIDEIILRLDGDVYEDNVPC"
    7439..8221
    /gene="MTH1101"

Query Match      100.0%; Score 10; DB 1; Length 10072;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTGAC 10
    |||||||||
Db 2658 TGTGTGTGAC 2649

RESULT 7
LOCUS AF006665/c
DEFINITION Bacillus subtilis 168 region at 182 mln containing the cge gene
ACCESSION AF006665
NID 52529445
VERSION AF006665.1 GI:2529445
KEYWORDS
SOURCE
ORGANISM
    . Bacillus subtilis.
    . Bacillus subtilis
    . Bacteria: Firmicutes: Bacillus/Clostridium group;
    . Bacillus/Staphylococcus group; Bacillus.
    1 (bases 1 to 30100)
    Ghim,S.-Y., Jeong,Y.-M., Choi,S.-K. and Park,S.-H.
    Sequence analysis of the 30 kb region (182%) of the Bacillus
    subtilis chromosome containing the cge cluster
    unpublished (1997)
    2 (bases 1 to 30100)
    Park,S.-H., Choi,S.-K., Jeong,Y.-M. and Ghim,S.-Y.
    Direct Submission
    Submitted (04-JUN-1997) Applied Microbiology Research Division,
    KIRIB, P.O.Box 115, Yusong-Gu, Taejeon 305-600, Korea
    location/Qualifiers
    1..30100
    /organism="Bacillus subtilis"
    /strain="168"
    /db_xref="taxon:1423"
    /map="182 mln"
    74..901
    /gene="yokB"
    74..901
    /gene="yokB"
    /gene="yokB"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /product="yokB"
    /protein_id="AB81138.1"
    /db_xref="PID:g2529446"

```



```
/db_xref="GI:2529446"
/translation="MKIDYSDLHINHWIMVNWQJIKWEKREIYNRLISNGEVL
VIAGDFTENQQTLMWLDEAAKQYERKVFYGNHDLYLKSKQKRYSDSLGRLNLI
OKADKNNVPLIKTETETKRYAGVPMYLPKGLGMDFFGKSDSNYILNKNYN
KYGVRAWMKESMDWYETLENTQVDFVSHVPHNHPSPPEPNTCYMDVPINAKH
WVGCHDLOAEFFKDGSTFHMNIGIPYIDNPVNVNIGEEVDSTKFFELKTEFI"
945..1136
/gene="yokC"
945..1136
/gene="yokC"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokC"
/protein_id="AAB81139.1"
/db_xref="PID:g2529447"
/db_xref="GI:2529447"
/translation="MENEVYVFCRCNHHLLFAKNPMTNLKVISEMDCPNCBEGYHN
WLSHGSEKREKENTMK"
1176..1307
/gene="yokD"
1176..1307
/gene="yokD"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokD"
/protein_id="AAB81140.1"
/db_xref="PID:g2529448"
/db_xref="GI:2529448"
/translation="WDSREQIDWTCNECNFSNIGNSDFSCPCDEIDIKPKNKILD"
1625..1891
/gene="yokE"
1625..1891
/gene="yokE"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokE"
/protein_id="AAB81141.1"
/db_xref="PID:g2529449"
/db_xref="GI:2529449"
/translation="WTFNIENSSDIHEFLIAAKNDVDPFLEWIKENGIPRLKNI
DFNKLIPNDRFLAMEFDEFLQNMILIKRFGVLFQOLIA"
2112..2333
/gene="yokF"
2112..2333
/gene="yokF"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokF"
/protein_id="AAB81142.1"
/db_xref="PID:g2529450"
/db_xref="GI:2529450"
/translation="MIEIFKDTGATHDLVYHSKINTFEWVDVEFDIVLSDSKELNKCYP
VKCENPYRINCKCFVASSIDIFSEVANS"
2539..2751
/gene="yokG"
2539..2751
/gene="yokG"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokG"
/protein_id="AAB81143.1"
/db_xref="PID:g2529451"
/db_xref="GI:2529451"
/translation="MCNRRNVITTYPEEDMSKYSILHQVGRIEVFQKEYSQYPMFARD
SEEDYNEKCLIMQDILKRNKKVSSPSF"
complement(3022..3303)
/gene="yokH"
complement(3022..3303)
```

```
/gene="yokH"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokH"
/protein_id="AAB81144.1"
/db_xref="PID:g2529452"
/db_xref="GI:2529452"
/translation="WTVLKSLEKVCYKKQMKPVDSPRLNLELMHEYSVSIPDLVECTG
LSKQRIINDYVGFGKSNMNIGMTFADAGCSIEELIYWNFKERROLTK"
3339..3926
/gene="yokI"
3339..3926
/gene="yokI"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokI"
/protein_id="AAB81145.1"
/db_xref="PID:g2529453"
/db_xref="GI:2529453"
/translation="WGANNQKVEANIEKSAADOKLFFRIRKDVNPMFLKGAAVSK
NKYDCLFFKGLYLPFPELSTYSKKSISFSEKIIKPOQIKYREATORPNIIPGLFOP
REPERKVFYVHINDELTKYIAKQIKHTYKNKYKASIPLAICEIECTEYRSKKKY
NTYVINKLCELKKSQSRDKPLHTNTPYKTYV"
3929..4105
/gene="yokJ"
3929..4105
/gene="yokJ"
/function="unknown"
/codon_start=1
/transl_table=11
/product="yokJ"
/protein_id="AAB81146.1"
/db_xref="PID:g2529454"
/db_xref="GI:2529454"
/translation="MEPYQREELKKTIVQKENSINITYTODEASNDLDEFYQOF
AQHLEALERKAE"
complement(4140..4565)
/gene="orfE5"
complement(4140..4565)
/note="orfE5: a potential ORF adjacent to the DNA
replication terminator Tery of B. subtilis; it includes
the right attachment site for phage Spbeta (4145-4160)"
/codon_start=1
/transl_table=11
/protein_id="AAB81147.1"
/db_xref="PID:g2529455"
/db_xref="GI:2529455"
/translation="MPKOQTAELKPFPHNKTLYVTGGTSGISQIVKRLMLTPKQVI
VFESKDSKOYVMSQKAYEDKRLFLVDGDRVQNVKMGVDIVFHARALKQVPICE
DHPEFAIQTNLIGQNVVEALSHRVQHYINISDKAYV"
4935..6083
/gene="yokK"
4935..6083
/gene="yokK"
/function="unknown"
/note="Orf181 at the 3' region of B. subtilis cgeAB"
/codon_start=1
/transl_table=11
/product="yokK"
/protein_id="AAB81148.1"
/db_xref="PID:g2529456"
/db_xref="GI:2529456"
/translation="MKVPEKTMLISTAGILSLTATSASAHYVNEBHHKVTAAHETD
PVASGDADDAIPAIVWHEKHPERSKITTNKSGLVYDLDKQKLSIEFGKLNWDL
RIDPPLNGEKIDIAASNSRSEGNKNTLEVAIIDGDKLKSIDPNHPISITNISEYGF
SLYHSQGTAFALVYKQGEFEQYEIVDGGKGYTKKVRREFKLNISQTEGLVADDEX
GNYIAEDEDAIWKNAEPGGSGQVVDRAAGDULIADIEGLTIYVAPNGKGYMAS
SOGNNSYAMYERQGRNRYVANFEITDGETSDGTDGIDVYFGFGIKRYPYIGITVAAQ
DGENIDNGOAVNQNFRIVSWEOIAQHLGEMPLDHRQVNPRLKDKSDSC"
complement(6154..7107)
```


CDS

/gene="cgeb"
/note="gene in the cge cluster, GerE-controlled genes"
complement(6154..7107)
/gene="cgeb"
/note="a protein in the cge cluster"
/codon_start=1
/transl_table=11
/product="Cgeb"
/protein_id="AAB81149.1"
/db_xref="PID:92529457"
/db_xref="GI:2529457"

very Match 100.0%; Score 10; DB 1; Length 30100;
est Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGTGTGTCAG 10
15427 TGTGTGTCAG 15418

RESULT 8
LOCUS AF015775
DEFINITION Bacillus subtilis yoda (yoda), yodB (yodB), yodC (yodC), yodD (yodD), ABC-transporter (yodE), permease (yodF), proteinase (ctpa), yodH (yodH), yodI (yodI), carboxypeptidase (yodJ), purine nucleoside phosphorylase (deod), yodL (yodL), yodM (yodM), yodN (yodN), yodO (yodO), yodP (yodP), acetylornithine deacetylase (argE), butyrate-acetoacetate CoA transferase (yodR), butyrate acetoacetate-CoA transferase (yodS), yodT (yodT), CgeE (CgeE), CgeD (CgeD), CgeC (CgeC), CgeA (CgeA), CgeB (CgeB), YzxA (YzxA), UDP-glucose epimerase (yodU), yodV (yodV), and yodW (yodW) genes, complete cds; and yodZ (yodZ) gene, partial cds.
AF015775
2415383
AF015775.1 GI:2415383

ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE

FEATURES
source

gene
CDS

1.29950
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
493..940
/gene="yoda"
493..882
/gene="yoda"

Unpublished (1997)
3 (bases 1 to 29950)
Wambutt, R., Lapidus, A., Sorokin, A. and Ehrlich, D.
Direct Submission
Submitted (23-JUL-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
Location/Qualifiers
1.29950

terminator
gene
CDS

/citation=[2]
/codon_start=1
/transl_table=11
/product="Yoda"
/protein_id="AAB72052.1"
/db_xref="PID:92415384"
/db_xref="GI:2415384"
/translation="MPLRFDVIEGRDEKSLQILLDTAQAQWAEARCEPERDQIVH
QHPANLLITDAGLGFORIKDMYITISRTAKTESOKELIALLARLERKEEISPD
LWVSTIENDGDAWMSFGLGEAOFNLNGL"
903..940
/gene="yoda"
complement(1289..1627)
/gene="yodB"
complement(1289..1627)
/note="similar to B. subtilis yodH protein"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YodB"
/protein_id="AAB72060.1"
/db_xref="PID:92415392"
/db_xref="GI:2415392"
/translation="MGNTMCPKKEASFSLGKRNGLIIRVIMDGPRKEITETIP
ISQMLAEKLELEQNEIYEROVLPETPVKVIITTEKGTALQAVFOEQAWADPCE
PDIVCEEEK"
1757..2407
/gene="yodC"
1757..2365
/gene="yodC"
/note="similar to E. coli oxygen-insensitive NAD(P)H
nitroreductase Nimb"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YodC"
/protein_id="AAB72053.1"
/db_xref="PID:92415385"
/db_xref="GI:2415385"
/translation="MTNLDVLRKASVREYDTPNPISEETLELLDTAKPAAVNL
OHMFTVEHSEDSKAEILPVAYNOKQIVSSAVVALIGLAKNENGEYPAELASGY
ITDEIKQTLIGINGAIOSEGTARSAFANSLAQAQNTIAKAGVDTCAIGENKE
QFQKQDISERTVPMLISIGKAVPAHQSNPLPSKYSTWL"
2373..2407
/gene="yodC"
complement(2408..3010)
/gene="yodD"
complement(2408..3010)
/note="similar to Leishmania mexicana CDG2-related protein
kinase CRK1"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YodD"
/protein_id="AAB72061.1"
/db_xref="PID:92415393"
/db_xref="GI:2415393"
/translation="MKHYEKTSNDVILLHGTGNGNHHDLISGRFTIDPDNHLIGVR
GSVLENGMPREFKRLSEGVFDEKLVVTRRELKPIDEAERHOFNRGRVAVGWSG
ANLASLFEHKDYDKGAILHHPVPIRGIELPDAGLPVFIAGKYPDLCKEESSE
LYRYRDSGASASVWODGGHOLTQHEAQAQREMYKAIIV"
complement(3026..3937)
/gene="yodE"
complement(3026..3937)
/note="similar to B. subtilis hypothetical ABC transporter
YybJ"
/citation=[2]
/codon_start=1
/transl_table=11
/product="ABC-transporter"

gene
CDS
/protein_id="AA072062.1"
/db_xref="PID:g2415394"
/db_xref="GI:2415394"
/translation="MKREGLHWTAFAADPOENRFTYEVIGLRLVYKTVNPDPCGY
HFRNGNGDPGCTMTTFPFQSGGTVGQGRVTFVSPGSLSWKRLKESGJS
LEKTLTGKGLFDJEDTDLPLAIMEADKSGKSWPDGTTTNEATGKMGVLLYSD
POATVILTTESFGYTKVAEDQIVRLASAVGVIVDVLHPEKRGVGGYVHVA
RTKKRQAKMLPIAENHLPSSEILDEYFTSVYREKGLLEFIANDDEPFMDERE
AEGTSKLPEMEKHKHQITDILPEL"
4321..5836
/gene="yodH"
4321..5811
/gene="yodH"
/note="similar to E.coli proline permease PutP"
/citation=[2]
/codon_start=1
/transl_table=11
/product="permease"
/protein_id="AA072054.1"
/db_xref="PID:g2415386"
/db_xref="GI:2415386"
/translation="MGNLTALITATITVIVTCVIGLAGRDSRPSVSEVSGGR
FGLVWFLVGAADLYATYFLGLSTAFPGSVAFRIPYSAIATFPLRLKRV
AKJHKLTLADYAKERNKSLASLVAIVGLMIPICQLSGIOTLQVAGGYTN
VKEVYIISFLVALYTFEFGIKGPTYAIIDLVVIMLVNVSFLHFNWGTPI
DLVKEAPQMLTIPSEGGKIPMTFTASISALAFMAHAAGVFAKGSADAVRKS
MFLVYIVLIVLIFGLFIAPLVLEEDTNRILHLHIONSYGVAGFAATIALAS
LFIQSIATGASNLFANNLRYDLHPVSGSKLITVRSVYVYIGALLFGLMFLPA
LVTLQLGVSQGVQIFPAIVASLFWKNOTKENTYTGLAGLAVTFYITQSAHGTE
GFWGLANMLAVVILNLEFVKNGSNVIEGLFKODANPNOKGA"
complement(5810..5836)
/gene="yodH"
complement(5840..7246)
/gene="ctpA"
complement(5846..7246)
/gene="ctpA"
/note="similar to Synecocystis sp. PCC6803
carboxy-terminal processing proteinase precursor"
/citation=[2]
/codon_start=1
/transl_table=11
/product="proteinase"
/protein_id="AA072063.1"
/db_xref="PID:g2415395"
/db_xref="GI:2415395"
/translation="MKRQLKLFVILITAVVASLITFTGNSIIGQKASAGDSKF
DKLKKATEQIKSDYQKTDKLVDAIKGMISLDIPYSTMDQEQAKSFEDTISAS
FEGIGAVERKDEILIVSIPKSPARKAKIPDQIKVNGRSKMGNNENAVALLR
GKGTGKYLELNAGVGNIDLSIKRDTIPVEYSEKDNNGEIIQTSSEETAKEL
TDAIDSLERKAGXYILDRNGPGLMEQALITVSNLEIDKGNIMOVEYNGSKRWK
AERKRYKPTVAVNDGTAASAEIYMAALHESNNGISETPGKCTVOTAKREYDQ
STVKLYAKMLTADGEMIRKKGIPOYKALIPDYAKIPIYADADTIKSGTGNNVKA
QMKLALGVKVNAMTDDFVSVQFQKREKLNTEGILLTGDTTKLMTELQKSLSD
NDTQMEKAIETLKKEM"
7399..8100
/gene="yodH"
7399..8100
/gene="yodH"
/note="similar to E.coli biotin biosynthesis protein BiotC"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yodH"
/protein_id="AA072055.1"
/db_xref="PID:g2415387"
/db_xref="GI:2415387"
/translation="MSRYLEMLSLFGVAGHPGLAFSKAVLOKAPSPDOPILNAGC
GTGQATVILHILYPTVVDKDPIMLEKAKRPNBGLAPAOELEHPPSESS
CVLSSEYLSRLSSLOEISRYLKPSCMLIGLEPAUKRPPAKRQKMDYGFCTC
HESEEMKILRSIFGQKTEAMSLPEMEPEPTTEMDLSOTIDPIYDITLQTHYQMLQ
LYSEYMGHCIFIAVK"
8188..8503
/gene="yodI"
gene

CDS
8188..8439
/gene="yodI"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yodI"
/protein_id="AA072056.1"
/db_xref="PID:g2415388"
/db_xref="GI:2415388"
/translation="MERYHLCKNHQKAVRTTERGRVHVRITRVRDRAVIAVWG
Query Match 100.0%; Score 10; DB 1; Length 29950;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGTGTGAC 10
Db 14707 TGTGTGTGAC 14716
RESULT 9
ASARGDG 2259 bp DNA BCT 16-FEB-1995
LOCUS Anabaena sp. (PCC 7120) argD gene.
ACCESSION X78854
NID 9474180
VERSION X78854.1 GI:474180
KEYWORDS argD gene; N-acetylornithine aminotransferase.
SOURCE Anabaena sp.
ORGANISM Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
REFERENCE 1 (bases 1 to 2259)
AUTHORS Flores, E.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1994) E. Flores, Instituto de Biología Vegetal,
Y Fitosintesis, Univ. de Sevilla-Csic, Facultad Biología, Apartado
1113, 41080 Sevilla, SPAIN
2 (bases 1 to 2259)
REFERENCE Floriano, B., Herrero, A. and Flores, E.
TITLE Analysis of expression of the argC and argD genes in the
JOURNAL Cyanobacterium Anabaena sp. strain PCC 7120
MEDLINE J. Bacteriol. 176 (20), 6397-6401 (1994)
FEATURES
source location/Qualifiers
1..2259
/organism="Anabaena sp."
/strain="PCC7120"
/db_xref="taxon:1167"
/clone_lib="wild type genomic in cosmid pDUCa"
/clone="pCSB335-2"
325..>1711
428..1711
/gene="argD"
428..1711
/gene="argD"
/codon_start=1
/transl_table=11
/product="N-acetylornithine aminotransferase"
/protein_id="CAA55410.1"
/db_xref="PID:g580729"
/db_xref="GI:580729"
/db_xref="SWISS-PROT:P54752"
/translation="MSLQTLIQATNPPESSGAASSPSTDSFSDASVNSTGYRPLAL
ERAGGRVMDTQGEYLDFAVAGIACITLHAPAVAVNTDIOKLHVSNNLYIPEQ
GELAGWIIQSCADRVFECNSGAENEAIKLARYAVRLDIEKPIITANASRGR
TLATITATGQATYKQYEDLPVGFHVNNVNDISAVEAATSEIDEDYRAAATLIEPLQ
GGGCVPRPVEYFKLRQICDPTGILINDEYQVGGSKRLMGTEIGVEPDITSA
KELGGGIPFGAMSKKFDVQFQGEHASTFGSNPACGVALAVCQTLERENLQNVQD
REBQUNSGRLAIARYPHRLTEVRWGLINLELADIPLTADYVAKAINEGLLVP
AGPKVVRVPLPVEAEIETALKLEKALATVIA"
1738..1816
/note="putative"
terminator
BASE COUNT 692 a 477 c 501 g 589 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 2259;
 Best Local Similarity 100.0%; Pred. No. 3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCAG 10
 |||||
 Db 1381 TGTGTGTCAG 1390

RESULT 10

BACWBA 5198 bp DNA BCT 02-FEB-1999
 LOCUS
 DEFINITION Bacillus subtilis modifier protein of major autolysin (cwba) gene
 and assumed lipoprotein (lppx) gene, complete cds.
 ACCESSION D10388
 GI216255

SYN D10388.1 GI:216255
 KEYWORDS lipoprotein; major autolysin.
 SOURCE Bacillus subtilis (strain:1685) DNA.
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
 Bacillus.

REFERENCE 1 (bases 1 to 5198)
 AUTHORS Sekiguchi, J.
 JOURNAL Direct Submission
 Submitted (29-JAN-1992) to the DDBJ/EMBL/GenBank databases. Junichi
 Sekiguchi, Textile Science and Technology, Shinsu University,
 Department of Applied Biology; 3-15-1 Tokida, Ueda-shi, Nagano 386,
 Japan (E-mail: jsekiguchi@ipc.shinsu-u.ac.jp, Tel:0268-21-5344,
 Fax:0268-21-5331)
 2 (sites)
 REFERENCES Kuroda, A. and Sekiguchi, J.
 JOURNAL Molecular cloning and sequencing of a major Bacillus subtilis
 autolysin gene
 J. Bacteriol. 173 (22), 7304-7312 (1991)
 3 (bases 1 to 5198)
 Kuroda, A., Rashid, M.H. and Sekiguchi, J.
 JOURNAL Molecular cloning and sequencing of the upstream region of the
 major Bacillus subtilis autolysin gene: a modifier protein
 exhibiting sequence homology to the major autolysin and the spoIID
 product
 J. Gen. Microbiol. 138 (Pt 6), 1067-1076 (1992)
 92407479
 Submitted (29-Jan-1992) to DDBJ by:
 Junichi Sekiguchi
 Department of Applied Biology
 Textile Science and Technology
 Shinsu University
 3-15-1 Tokida, Ueda-shi
 Nagano 386
 Japan
 Phone: 0268-22-1215
 Fax: 0268-22-4079.
 Location/Qualifiers
 1. 5198
 /organism="Bacillus subtilis"
 /strain="1685"
 /db_xref="taxon:1423"
 908..911
 927..934
 943..949
 955..1263
 /gene="lppx"
 955..1263
 /gene="lppx"
 /codon_start=1
 /transl_table=11
 /product="lipoprotein"
 /protein_id="BA01223.1"

REFERENCE 1 (bases 1 to 5198)
 AUTHORS Sekiguchi, J.
 JOURNAL Direct Submission
 Submitted (29-JAN-1992) to the DDBJ/EMBL/GenBank databases. Junichi
 Sekiguchi, Textile Science and Technology, Shinsu University,
 Department of Applied Biology; 3-15-1 Tokida, Ueda-shi, Nagano 386,
 Japan (E-mail: jsekiguchi@ipc.shinsu-u.ac.jp, Tel:0268-21-5344,
 Fax:0268-21-5331)
 2 (sites)
 REFERENCES Kuroda, A. and Sekiguchi, J.
 JOURNAL Molecular cloning and sequencing of a major Bacillus subtilis
 autolysin gene
 J. Bacteriol. 173 (22), 7304-7312 (1991)
 3 (bases 1 to 5198)
 Kuroda, A., Rashid, M.H. and Sekiguchi, J.
 JOURNAL Molecular cloning and sequencing of the upstream region of the
 major Bacillus subtilis autolysin gene: a modifier protein
 exhibiting sequence homology to the major autolysin and the spoIID
 product
 J. Gen. Microbiol. 138 (Pt 6), 1067-1076 (1992)
 92407479
 Submitted (29-Jan-1992) to DDBJ by:
 Junichi Sekiguchi
 Department of Applied Biology
 Textile Science and Technology
 Shinsu University
 3-15-1 Tokida, Ueda-shi
 Nagano 386
 Japan
 Phone: 0268-22-1215
 Fax: 0268-22-4079.
 Location/Qualifiers
 1. 5198
 /organism="Bacillus subtilis"
 /strain="1685"
 /db_xref="taxon:1423"
 908..911
 927..934
 943..949
 955..1263
 /gene="lppx"
 955..1263
 /gene="lppx"
 /codon_start=1
 /transl_table=11
 /product="lipoprotein"
 /protein_id="BA01223.1"

Faculty of

FEATURES
 SOURCE

-35-signal
 -10-signal
 RBS
 gene
 CDS

sig_peptide
 mat_peptide
 RBS
 gene
 sig_peptide
 CDS

mat_peptide

gene

CDS

BASE COUNT 1750 a
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 5198;
 Best Local Similarity 100.0%; Pred. No. 2.6e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCAG 10
 |||||
 Db 5151 TGTGTGTCAG 5160

/db_xref="PID:d1001693"
 /db_xref="PID:g216256"
 /db_xref="GI:216256"
 /translation="MKFIALLEPFIILSGCCGVNSQSGEDVSPDSNIETKEGTG
 IADHTIEVTVNDPEPSLDITEESTSDLRKNSGDKVTIYERKNDGOLLKDIERN
 "

955..1002
 /gene="lppx"
 1003..1260
 /gene="lppx"
 /product="lipoprotein"
 1269..1275
 1287..13404
 /gene="cwba"
 <1287..1361
 /gene="cwba"
 <1287..13404
 /gene="cwba"
 /codon_start=1
 /transl_table=11
 /product="modifier protein of major autolysin"
 /protein_id="BA01224.1"
 /db_xref="PID:d1001694"
 /db_xref="PID:g2160197"
 /db_xref="GI:2160197"
 /translation="LNSCKOLIVCSLAAILLIPSVFAADSNISVKLNYIGNKSSI
 SLSPGTGYKVTGDNVAATDRFAGASRYETATLANSQOMKNTVILVNDIFDALPV
 IPLAKINAPLVEFTOPDTLTTERQIAKPNPNILIGARSISQEKRLKSGAV
 KRISGRNVYLVENIKAKOMGSYDKAIIVTGRVODALAIAPAAHGYPIILREKDL
 VMDPLDPAOVLITIGSSPSVSDVNOIKKTGTORIGSTREYETANTIKLKADYV
 VMTNGTAYAVLIGASLASAKNSKQIFVKKDSTYPAAKSITKDKATAYDFIGSISL
 SAEVNSLADEFLADGTYNLKINSKRLKLVNKTGYNSLRKPEVYSNLSLIG
 KOYLGVNESIESTKYIRPVENIPEEDYIKGYIPNMPASWSLEALKAQTVARYS
 ITRKGTVPPTAFQVYGGYSNNSNTKAVEQTKYKLYGSLITAYSSNGGYTE
 ASNEVSSVPLVIAKKDTPQIGTILSLKOLDKSLDLTPSSWSSAETGSA
 RLGGVKNMILKNETSADSVKIASIDLSFGCTGCGRATASMKYKPYFKSSGTGSN
 LSKITITISVPTSELRTMGATVFKSTIVYKKDTSKITISGKGIGHSQYGAAR
 AERGDSSYLKFFYPETLTYS"
 1362..13401
 /gene="cwba"
 /product="modifier protein of major autolysin"
 3443..4933
 /gene="cw1B"
 <3443..4933
 /gene="cw1B"
 /EC_number="3.5.1.28"
 /codon_start=1
 /transl_table=11
 /product="N-acetylmuramoyl-L-alanine amidase"
 /protein_id="BA01225.1"
 /db_xref="PID:d1001695"
 /db_xref="PID:g2160198"
 /db_xref="GI:2160198"
 /translation="LSRYIVLMCEGLILFPPTALADNSVRVSGSNNTYGAOIS
 KQMTSTASTAVIVGSSYADPAISAPLAKOKNAPLITNSDKLSYETKRLKEMQKN
 VILVGGTIPAVSNTANOIKSLGISIRIAGSNRYDPAARAKMGATSKAVILNGELY
 ADAPAVIYPAKNGYPILEFNKTSINSATSVYKDGISSTVVGATGSGISNTYKLN
 PSPTRISGSRNRYLANIVOKLNSTVYVSGFSYSPDSIAGATIAAKKOSLIITLN
 GENLSTGARKVIGSKSNMSNMIIGTNPASTYKVALOKNPNVGETEIFIDGDDOOSG
 AIGNGLERKVNIDIAKRVNTRKLNAGALPYLRSNDTRISLOERPNKASADADLEFL
 SIHANNDSSSPGSEETIDTYOANNSKRLAIOIPKLANLGTDRGVKTAIFYVI
 KYSKMPSVLEIVNFTINASDASKLOAVYKDKRAQAQIHQTVSYRR"

RESULT	11
LOCUS	BACCMWL
DEFINITION	Bacillus subtilis N-acetylmuramoyl-L-alanine amidase (cwlB) gene,
ACCESSION	M81324 M61747
NID	9142805
KEYWORDS	M81324.1 GI:142805
SOURCE	N-acetylmuramoyl-L-alanine amidase; cwlB gene; major autolysin.
ORGANISM	Bacillus subtilis (strain 168) DNA.
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;
AUTHORS	Bacillus/staphylococcus group; Bacillus.
TITLE	1 (bases 1 to 2497)
	Kuroda,A. and Sekiguchi,J.
	Molecular cloning and sequencing of a major Bacillus subtilis
	autolysin gene
JOURNAL	J. Bacteriol. 173, 7304-7312 (1991)
DLINE	92041629
FEATURES	
source	Location/Qualifiers
	1..2497
	/organism="Bacillus subtilis"
	/strain="168S"
	/db_xref="taxon:1423"
	730..734
RBS	/gene="cwlB"
	730..2269
gene	/gene="cwlB"
	742..2232
CDS	/gene="cwlB"
	/EC_number="3.5.1.28"
	/codon_start=1
	/transl_table=1
	/product="N-acetylmuramoyl-L-alanine amidase"
	/protein_id="AA02371.1"
	/db_xref="pid:g142805"
	/db_xref="GI:142806"
	/translation="MRSIKVLYMCFLGLIFVPALADNSVKRVGSSNRGTAVQIST
	KOMSTASTAVIVGGSSYADAISAAPLQKRNAPLLTNSDKLSYEKTRUKENQTKNK
	VIIIGTFPVAVSNTANOIKSLGISIRLAGSNRPDIARVAKAGATSKAVILNGFDF
	ADAPRIVPAKNKGPILEFTKTGSINSATTSVIDKDGISSVVVGSGISINTVYNKL
	PSPRIGSSNRVELAANTVOKLNTSTNVVYSSNGFSYDSDTAGATTAAKKKOSLIITNN
	GENSITGRKITGSKNSNFHITGNTPAVSRKYANQLKNPYGETIIFDPHGQDDSG
	AIGGLEKEVNLDIAKRVNTRLKNSGLPVLSNSDTFYSLOERVNFAASAQDLFF
	SIHANADSSPSNGSEITYDTTOYANASKRLAEIOIQLPLANLTGRDGVXTAAFYHY
	KYSKMPSVLVETAFITNASDAKLUOAYVKDKAAQAIDHGTVSYYR"
	2236..2269
terminator	/gene="cwlB"
	/note="ho-independent terminator"
BASE COUNT	876 a 442 c 508 g 671 t
ORIGIN	
Query Match	100.0%; Score 10; DB 1; Length 2497;
Best Local Similarity	100.0%; Pred. No. 3e+03;
Matches	10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 TGTGTCTAG 10
DB	2450 TGTGTCTAG 2459
RESULT	12
LOCUS	BACENZYM1
DEFINITION	Bacillus subtilis 168 enzyme I of the PEP:phototransferase system
ACCESSION	L08809
NID	9289267
VERSION	L08809.1 GI:289267
KEYWORDS	PEP:phototransferase; enzyme I; lyase; spore photoproduct lyase.

```

SOURCE
ORGANISM      Bacillus subtilis (strain 168) DNA.
REFERENCE     Bacillus subtilis
AUTHORS       Bacteria: Firmicutes; Bacillus/Clostridium group;
              Bacillus/Staphylococcus group; Bacillus.
TITLE         1 (bases 1 to 2277)
              Fajardo-Cavaasos, P., Salazar, C. and Nicholson, W.L.
              Molecular cloning and characterization of the Bacillus subtilis
              spore photoprotectant lyase (spl) gene, which is involved in repair of
              UV radiation-induced DNA damage during spore germination
JOURNAL       J. Bacteriol. 175, 1735-1744 (1993)
MEDLINE
FEATURES
Source        Location/Qualifiers
              1..2277
              /organism="Bacillus subtilis"
              /strain="168"
              /db_xref="taxon:1423"
              <1..522
              /note="Phototransferase system"
              /codon_start=1
              /transl_table=1
              /product="enzyme 1"
              /protein_id="AA22414.1"
              /db_xref="PID:g289268"
              /db_xref="GI:289268"
              /translation="EFKAKAILLEEKKEKLVKAGQAVSDIEVGWMMVEIPSTAYIADQ
              FAKVDPEFSGTNDLIQIYTAADPMRNRVSYLYQPYNAILRLIYIEAAHKQKAW
              GWCQMGADETAIPILGLGDFSMSTSTILPARTQISLKQDAESFEKILSMST
              TEVVAFFVKETFK"
              613..852
              /codon_start=1
              /transl_table=1
              /function="unknown"
              /product="9000 kDa protein"
              /protein_id="AA22415.1"
              /db_xref="PID:g289269"
              /db_xref="GI:289269"
              /translation="MSNQFNAGDTVVYIYRPHANAHAKIETIVHHPYHEGSLSE
              IYETHPFAEDDAVFASIEAKSLYKELFDIPE"
              930..1958
              /codon_start=1
              /transl_table=1
              /product="spore photoprotectant lyase"
              /protein_id="AA22416.1"
              /db_xref="PID:g289270"
              /db_xref="GI:289270"
              /translation="MQNPFPQVLYIEPRALYELPQLODKPFENMGIEIRETTSNHC
              VRNIPKGNHLOQVYNAKSLTLYIGRTILKFDSSKPSAEYAIPEPAGCGHCHYCLQI
              TMSGKPYIRTYVYNEELLDDQDKYMKRAPERFEEASCTSDIYGDILHTTLKRAIE
              HFGSDSGKLRFYTKRFHVDHLDAKANGKTRFRFSTINADYVINKPEGTSPLDKRIT
              KAAVAVANAGIPPLGFIIVAPILYIHGEMEGSYRLFEKLDNALPDQVRHDTFVLQHRF
              KPAARVIEKNPKTKLELDDEKRRIGKRGYIGYIYQKDEHALREALEYIDTFE
              NAKIEYFT"
BASE COUNT    736 a      470 c      494 g      577 t
ORIGIN
Query Match   100.0%; Score 10; DB 1; Length 2277;
Best Local Similarity 100.0%; Freq. No. 3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY            1 TGTGTGTCAG 10
              |||||
              |||||
Db            1389 TGTGTGTCAG 1380
RESULT 13
LOCUS       BSU55043      8397 bp      DNA      BCT      09-MAY-1996
DEFINITION  Bacillus subtilis plasmid pPOD2000 Rep, RapAB, RapA, ParA, ParB,
              and ParC genes, complete cds.
ACCESSION   U55043
            g1305508
VERSION     U55043.1      GI:1305508

```



```

KEYWORDS      Bacillus subtilis.
SOURCE         Bacillus subtilis.
ORGANISM       Bacteria; Firmicutes; Bacillus/Clostridium group;
                Bacillus/staphylococcus group; Bacillus.
REFERENCE      1 (bases 1 to 8397)
                Thorsted,P.B., Rowlinson,C., Gleave,A.P. and Thomas,C.M.
                A putative serine protease helps to ensure stable inheritance of
                cryptic plasmid pPOD2000 from Bacillus subtilis
                Unpublished (1996)
                2 (bases 1 to 8397)
                Thomas,C.M.
                Direct Submission
                Submitted (15-APR-1996) Christopher M. Thomas, School of Biological
                Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT,
                UK

FEATURES
  source        Location/Qualifiers
                1. .8397
                /organism="Bacillus subtilis"
                /plasmid="pPOD2000"
                /db_xref="taxon:1423"
                complement(353. .384)
                complement(393. .917)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09148.1"
                /db_xref="PID:g1305509"
                /db_xref="GI:1305509"
                /translation="MKKPKKENVPLSLVYRAVAFQPISSAKAVESEGEFVNK
                QLDALAEBSDLTQRTIVKDGILYMKRDKSKAKFRRSQSLVDVYNSINGK
                IVAKGSIINDVLTNHPVKDKVSSKVDVDRGELSTKIVASGYVLMSNETNAIKI
                VAAAGALASLRHE"
                complement(822. .917)
                /note="ORF1"
                complement(923. .926)
                /note="ORF1"
                /note="ORF1"
                1042. .1048
                /note="ORF2"
                1056. .1112
                /note="ORF2"
                1056. .1376
                /note="ORF2"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09149.1"
                /db_xref="PID:g1305510"
                /db_xref="GI:1305510"
                /translation="MKNIPTIVATITPFSVSLFENSILFVSVLSIPFILLNF
                VKETNKVKQLKVLINILAGSTIISIFILENIYISGILLGITSVIITISIIK
                AGYN"
                complement(1362. .1381)
                complement(1383. .1580)
                /note="ORF3"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09150.1"
                /db_xref="PID:g1305511"
                /db_xref="GI:1305511"
                /translation="MARIEGIVFPCPCNTRFDMQHKVSSRYCKCKKTNLKTKLLT
                MDNEFHGQFCCKCKEMTGEY"
                complement(1369. .1592)
                /note="ORF3"
                complement(1591. .2706)
                /gene="rep"
                complement(1591. .2517)
                /note="rep"
                /note="replication initiator protein; similar to other
                plasmids"
                /codon_start=1
                /transl_table=11
                /product="Rep"

  terminator    1. .8397
                /organism="Bacillus subtilis"
                /plasmid="pPOD2000"
                /db_xref="taxon:1423"
                complement(353. .384)
                complement(393. .917)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09148.1"
                /db_xref="PID:g1305509"
                /db_xref="GI:1305509"
                /translation="MKKPKKENVPLSLVYRAVAFQPISSAKAVESEGEFVNK
                QLDALAEBSDLTQRTIVKDGILYMKRDKSKAKFRRSQSLVDVYNSINGK
                IVAKGSIINDVLTNHPVKDKVSSKVDVDRGELSTKIVASGYVLMSNETNAIKI
                VAAAGALASLRHE"
                complement(822. .917)
                /note="ORF1"
                complement(923. .926)
                /note="ORF1"
                /note="ORF1"
                1042. .1048
                /note="ORF2"
                1056. .1112
                /note="ORF2"
                1056. .1376
                /note="ORF2"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09149.1"
                /db_xref="PID:g1305510"
                /db_xref="GI:1305510"
                /translation="MKNIPTIVATITPFSVSLFENSILFVSVLSIPFILLNF
                VKETNKVKQLKVLINILAGSTIISIFILENIYISGILLGITSVIITISIIK
                AGYN"
                complement(1362. .1381)
                complement(1383. .1580)
                /note="ORF3"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09150.1"
                /db_xref="PID:g1305511"
                /db_xref="GI:1305511"
                /translation="MARIEGIVFPCPCNTRFDMQHKVSSRYCKCKKTNLKTKLLT
                MDNEFHGQFCCKCKEMTGEY"
                complement(1369. .1592)
                /note="ORF3"
                complement(1591. .2706)
                /gene="rep"
                complement(1591. .2517)
                /note="rep"
                /note="replication initiator protein; similar to other
                plasmids"
                /codon_start=1
                /transl_table=11
                /product="Rep"

  RBS           -10_signal
                /protein_id="AA09151.1"
                /db_xref="PID:g1305512"
                /db_xref="GI:1305512"
                /translation="MAAHYDGLERKTPGAPYGGKAEKVCNCACALVERDPTETGRLLT
                SNVYCKVRKSCMCAKRSRSLKIAHKHKLVEEANKREKPEWITLTLVRVDDGDLKO
                TTDMMKAMNRLFGYKVKTSVSGYFRALETINHDONTLHPHPHVLIPKRSYFCK
                RYRKQAWSTLKRRAKLDITPYVHQVAGKKGIDEGIEKREKMEQKALIEIS
                KYPVKDITVRGNEVEENLDTVYLLDDLANRRLIGYGGILKEIKRELNLTDABEDG
                LVLEEDDVAVAETFEVMAHVMVGIKNTYIIQ"
                complement(2524. .2531)
                /gene="rep"
                complement(2677. .2682)
                /gene="rep"
                complement(2701. .2706)
                /gene="rep"
                complement(2825. .2842)
                /standard_name="transcription terminator"
                complement(3293. .3424)
                /gene="rapAB"
                complement(3293. .3409)
                /gene="rapAB"
                /note="gene product with similarity to the rapAB gene
                product downstream of rapA on the Bacillus chromosome and
                rap40 and rap60 gene products from plasmids pTA1040 and
                pTA1060 respectively"
                /codon_start=1
                /transl_table=11
                /product="RapAB"
                /protein_id="AA09152.1"
                /db_xref="PID:g1305513"
                /db_xref="GI:1305513"
                /translation="MTFKIMAVLLIATVAVPVGSLATODNSVVASRRAT"
                complement(3339. .4535)
                /gene="rapA"
                complement(3339. .4517)
                /gene="rapA"
                /note="similar to the rapA gene product (formerly gsa) on
                the Bacillus chromosome and rap40 and rap60 gene products
                from plasmids pTA1040 and pTA1060, respectively"
                /codon_start=1
                /transl_table=11
                /product="RapA"
                /protein_id="AA09153.1"
                /db_xref="PID:g1305514"
                /db_xref="GI:1305514"
                /translation="MGTIAPVVGOKINEMVRIKRLNVTDAELINEVKNELDME
                DEQVLYXFOLEMRHOMLEVNPSTNTYKADYLAVBEGKRLGIMGYINFRG
                MFTSGEITKATITTRIEKRLDKVADLEERAEFFKMAEITLHKQTHLSYTYEQ
                AHAYKANSIYQVRLIQCPVANGVDDLVSYKRALPHKRALQESIALGHADITTKG
                LNLGNCFPNODVAVATAYYQALDELGEIKAEVVOACYDAIITFFGQKREELD
                YCVKGSYAEKYDSDMEVDENLHELIVTSPSVENOKFDQLVSRGPLYEDIAL
                EAGRFYNDIERLDSVFYFKMIOVKLQORDGEVDV"
                complement(3417. .3424)
                /gene="rapAB"
                complement(4527. .4535)
                /gene="rapA"
                complement(4587. .4625)
                /note="ComAP recognition site"
                complement(5088. .5297)
                /note="single strand replication origin; this is the site
                that lagging strand replication is initiated on the
                plasmid during rolling circle replication"
                5930. .5937
                /note="ORF4"
                5946. .6278
                /note="ORF4"
                /codon_start=1
                /transl_table=11
                /protein_id="AA09154.1"
                /db_xref="PID:g1305515"
                /db_xref="GI:1305515"
                /translation="KDNQETWHSIKLEKLEQIDPTTLRRYTRQGHHLNKKNGSY
                LVANESIPLLKRIKRLYVEGSKIENIETTLTNEGHPITITIDNHSTALTINESIED

```


Azavedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
 Borris, R., Boutsier, L., Brans, A., Brann, M., Brinell, S.C.,
 Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
 Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J.,
 Daler, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
 Emerson, P.T., Entian, K.D., Eriington, J., Fabret, C., Ferrati, E.,
 Follger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
 Gallon, N., Ghim, S.T., Glaser, P., Gottreau, A., Golligly, E.J.,
 Grand, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
 Henaud, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
 Itaya, M., Jones, L.J., Joris, B., Karamata, D., Kasahara, Y.,
 Klier, Blanchard, M., Klein, C., Kobayashi, Y., Koelter, P.,
 Konigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
 Ladinolis, S., Lauder, U., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
 Masuda, S., Manuel, C., Medigue, C., Medina, N., Mellado, R.P.,
 Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
 Owawa, K., Ogimura, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
 Portetlelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
 Punnett, B., Rapoport, G., Ray, M., Reynolds, S., Rieger, M.,
 Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, I., Sato, T.,
 Scanlan, E., Schleich, S., Schroeder, R., Scottone, F., Sekiguchi, J.,
 Sekowska, A., Serot, S.J., Serro, P., Shin, B.S., Solido, B.,
 Sorocini, A., Taccioni, E., Takagi, T., Takahashi, H., Takemaru, K.,
 Takeuchi, M., Takaoishi, A., Tanaka, T., Terstra, P., Tojima, A.,
 Tosato, V., Uchlyama, S., Vandenberg, M., Vannier, F., Vassart, A.,
 Viari, A., Wachi, R., Wedler, E., Wedler, H., Weitenegger, T.,
 Winters, P., Wipet, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
 Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
 Danchin, A.
 The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
 Nature 390 (6657), 249-256 (1997)
 98044033
 2 (bases 1 to 213190)
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 Direct Submission
 Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
 Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
 68 89 48
 Location/Qualifiers
 1. 213190
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 33. 740
 /gene="ydfS"
 33. 740
 /gene="ydfS"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12373.1"
 /db_xref="PID:e1182533"
 /db_xref="PID:g2632867"
 /db_xref="GI:2632867"
 /db_xref="SPRMBL:P96697"
 /translation="MTELEAVIRATVASEPILLIARILIKONPISQNTITDFIAATIG
 AIAAGAAVNTSIRKPHMAISFSLFVITLILFSLSKKKLRFRRAGDFTVLONSKI
 LSNKKRATIDLYNOOLREKEITENIEVLALETNGQLVLRKPOFRHTKQDLM
 IYVNDQRPILINDGELIENNLKONRLTESWMLTELRKIDIKVETIVAVILNGSD
 IYVDQKXDIISVPMKRE"
 complement(755. 1186)
 /gene="ydfT"
 complement(755. 1186)
 /gene="ydfT"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB12374.1"
 /db_xref="PID:e1182534"

[illegible]

CDS
/gene="yfmr"
6245..8134
/gene="yfmr"
/function="unknown"
/note="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB12566.1"
/db_xref="PID:e1182727"
/db_xref="PID:g2633061"
/db_xref="GI:2633061"
/db_xref="SPTREMBL:006476"

/translation="MSILKAKENIKKTYGDKTLFDHISFHIENERIGLIPNGTGST
LKVVIAGLESIEEGEITKSGSVQVELHODPELPAGQTVLEHIYSGESAVAKTLREYE
KALVELGKDPENEDROKHLAAQAKMDANNAMDANTLAKTVLSKLGNDVTKPVNELS
GGOKKRVIAIKNLIDPADLLIDEPNHLNETIEMLEGYLSQPGAVMLVTHDRFEL
NRYTNRIYELERGSITYKGNVEVLEKRAEREOAEOKETRONLLRELAWLRGA
KARSTOKARIDREVTKEOTGPOSSGSLDPAIGSHRIGKOVTEAEVMTAYDGRMLY
DRNELVTPGERIGIIGPNGIGKTTLLNALAGKHTPDGDIITGQTVRIQYIYQDHS
MNGELKVIDYIKETAEVKTAQDMITAEQMLERFEPKSMQYIIRKISGGERRLY
LLOVLMQEPNVFLDEPTNDIDETLSVLEDYIDQEPGVITVSHDRYFLDRVVDRLI
VFQNGVYSRFQGSYSDYMESEKAKKAPPAEAEKTAEEAPKKKKRLSYKQDLEMD
GIEDKIAQLEKHEQLEADIAAGSDGKIQELMAEQAKTAEELEAAMDRTLSIMI
EELSES"

terminator

8138..8159
/gene="yfmr"
/gene="yfmr"

gene

8257..8740
/gene="yfmQ"
/gene="yfmQ"

CDS

8257..8703
/gene="yfmQ"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12567.1"

/transl_table=11
/protein_id="CAB12567.1"

Query Match

100.0%; Score 10; DB 1; Length 208430;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTCAG 10
|||||

DB 8214 TGTGTGTCAG 8223

Search completed: September 17, 1999, 22:01:06
Time: 16249 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:30 ; Search time 425.19 Seconds

(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518A-87

Perfect score: 10

Sequence: 1 TGTGTCTCAG 10

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

abase: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	100.0	7377	1 003005	Full length T4 CDN
2	100.0	6151	1 003006	Full length T4 CDN
3	100.0	595	1 N90696	DNA sequence of po
4	100.0	591	1 N91782	DNA probe compleme
5	100.0	1900	1 N92624	Nucleotide sequence
6	100.0	5518	1 N90647	Nucleotide sequence
7	100.0	5565	1 N90648	Nucleotide sequence
8	100.0	5413	1 N90649	Nucleotide sequence
9	100.0	5365	1 N90646	Nucleotide sequence
10	100.0	1834	1 004690	Encodes Mammalian
11	100.0	7316	1 004555	SV40 initial promo
12	100.0	828	1 N81632	Human Bone Morphog
13	100.0	1605	1 N80632	Plasmid p170-2 enc
14	100.0	7533	1 N81970	Human Bone Morphog
15	100.0	600	1 N80979	Plasmid pEMPI-CPA.
16	100.0	533	1 N80239	Sequence of SV40 e
17	100.0	664	1 005513	Region between loc
18	100.0	646	1 005516	Upstream sequence
19	100.0	645	1 005517	Upstream sequence
20	100.0	641	1 005515	Upstream sequence
21	100.0	664	1 005514	Upstream sequence
22	100.0	6414	1 006283	Upstream sequence
23	100.0	7377	1 005607	Plasmid p170-2 inc
24	100.0	6151	1 005608	Plasmid p170-2 inc
25	100.0	560	1 006235	Plasmid pBG381 inc
26	100.0	455	1 N70071	Bovine male-specific
27	100.0	2400	1 N70061	Puili-BgIII fragme
28	100.0	5522	1 N70062	5' region of human
29	100.0	828	1 N70358	Smu3-EcoRI 5.4 kb
30	100.0	7533	1 N70491	Sequence of pSV-G
31	100.0	2659	1 N70216	Entire sequence of
32	100.0	6151	1 N90644	Extracellular alka
33	100.0	6149	1 N90645	Sequence of T4 lym
34	100.0	560	1 N60553	Sequence of T4 lym
35	100.0	1607	1 014036	Fragment of plasmid
36	100.0	1829	1 014627	Human BMP-2A in la
37	100.0	6051	1 014934	Human Glioblastoma
38	100.0	2721	1 N50114	pBG381 sequence. N
39	100.0	782	1 N50388	DNA sequence encod
40	100.0	840	1 N50390	Bovine interferon
41	100.0	840	1 N50391	Bovine interferon
42	100.0	840	1 N50391	Bovine interferon
43	100.0	6165	1 020324	sol. thesus CD4 seq

ALIGNMENTS

44 10 100.0 595 1 N81528 Sequence of a port
c 45 10 100.0 638 1 X40032 Prostate cancer as

RESULT 1
ID 003005 standard; CDNA; 7377 BP.
AC 003005;
DE 29-MAY-1990 (first entry)
DE Full length T4 CDNA of plasmid p170-2.
KW Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.
PN W08911860-A.
PD 14-DEC-1989.
PE 08-JUN-1989; U02453.
PR 10-JUN-1988; US-204645.
PR 20-APR-1989; US-341080.
PA (GEO-) General Hospital Corp.
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
DR WPI: 90-007302/01.
DR P-PSDB; R04031.
PT Combinations of soluble T4 protein and anti-retroviral agent -
PT having synergistic activity in treatment and prevention of AIDS,
PS Disclosure: h.
CC Soluble T4 constructs may be produced by truncating this sequence, to
CC remove the transmembrane and intracytoplasmic domains whilst retaining
CC the extracellular region responsible for HIV binding. The sol. T4 is
CC combined with an anti-viral agent such as AZT.
SQ See also 003006.
SQ Sequence 7377 BP; 1760 A; 1954 C; 1902 G; 1761 T;

Query Match 100.0%; Score 10; DB 1; Length 7377;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10
DB 22 TGTGTCTCAG 31

RESULT 2
ID 003006 standard; CDNA; 6151 BP.
AC 003006;
DE 29-MAY-1990 (first entry)
DE Full length T4 CDNA of plasmid pBG381.
KW Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.
PN W08911860-A.
PD 14-DEC-1989.
PE 08-JUN-1989; U02453.
PR 10-JUN-1988; US-204645.
PR 20-APR-1989; US-341080.
PA (GEO-) General Hospital Corp. (BIOJ) Biogen Inc.
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
DR WPI: 90-007302/01.
DR P-PSDB; R04032.
PT Combinations of soluble T4 protein and anti-retroviral agent -
PT having synergistic activity in treatment and prevention of AIDS,
PS Disclosure: h; 2; 100pp; English.
CC Soluble T4 constructs may be produced by truncating this sequence, to
CC remove the transmembrane and intracytoplasmic domains whilst retaining
CC the extracellular region responsible for HIV binding. The sol. T4 is
CC combined with an anti-viral agent such as AZT.
SQ See also 003006.
SQ Sequence 6151 BP; 1493 A; 1589 C; 1586 G; 1483 T;

Query Match 100.0%; Score 10; DB 1; Length 6151;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10
 |||||
 DB 26 TGTGTCTCAG 35

RESULT 3

ID N90696 standard; DNA; 595 BP.

AC N90696;

DE 15-JUL-1990 (first entry)

DE DNA sequence of portion of plasmid pF8CIS containing SV40 enhancer and

DE promoter, cytomegalovirus splice donor site and Ig intron

KW plasmid pF8CIS; cytomegalovirus; SV40; enhancer; promoter;

KM stabilising sequence; splice donor intron sequence; Ig region;

KM splice acceptor sequence; human embryonic kidney cells (293); JMW2.

Key Cytomegalovirus; SV40; +

Promoter

Location/Qualifiers
 1..360
 /tag= a
 /note="SV40 enhancer and promoter"

misc_rna

361..580
 /tag= b
 /note="Stabilising sequence which includes

cytomegalovirus donor and intron sequence, the Ig

variable region intron and splice acceptor sequence"

cds

480..580
 /tag= c
 /note="rge synthetic 100mer"

misc_feature

573..590
 /tag= d
 /note="psi-clai converter"

EP-309235-A.

29-MAR-1989.

22-SEP-1988; 308784.

25-SEP-1987; US-101712.

(GETH) Genentech Inc.

Gorman CM;

WPI: 89-095394/13.

Transient expression system for recombinant proteins -

comprising eukaryotic host cell transfected with vector

encoding trans-activating protein and expression vector

Disclosure: 44 pp; English.

Expression vector pF8CIS contains the SV40 enhancer and promoter, the

cytomegalovirus splice donor site and a portion of the intron, the Ig

intron and splice acceptor site, the cDNA encoding factor VIII and the

SV40 polyadenylation and transcription termination sites.

It is used as a method for prodn. of a

desired heterologous protein in a eukaryotic host cell.

The method comprises: transfecting a eukaryotic host cell with a vector

encoding trans-acting protein; transfecting the host cell with

an expression vector having a sequence of a ds DNA comprising, a

stabilising sequence downstream of a promoter and upstream of a DNA

encoding the amino acid sequence of the desired protein, and a poly-

adenylation sequence upstream to a transcription termination site.

The pref. host cells are human embryonic kidney cells (293) or JMW2.

The method provides useful quantities of a desired protein in a

relatively short period of time without having to establish continuous

prodn. The transient expression system optimises the interaction between

specific vector components and certain trans-activating proteins.

Sequence 595 BP; 136 A; 172 C; 146 G; 141 T;

Query Match 100.0%; Score 10; DB 1; Length 595;

Best Local Similarity 100.0%; Pred. No. 6e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10
 |||||
 DB 18 TGTGTCTCAG 27

RESULT 4

ID N91782 standard; DNA; 591 BP.

AC N91782;

DE 16-MAR-1990 (first entry)

DE DNA probe complementary to human papilloma virus type 11

KW Cervical cancer.

OS Human papilloma virus type 11.

PN WO8909940-A.

PD 19-OCT-1989.

PE 04-APR-1989; U001318.

PR 04-APR-1988; US-177404.

PI George AL, Groff DE;

(ONCO) Oncor Inc.

WPI: 89-324314/44.

PT Rapid detection of specific human papilloma virus genotypes + by

hybridisation of DNA digest with new labelled nucleic acid probes

PS Claim 34; page 44; 81pp; English.

CC Obd. by cutting HPV11 with BamHI and NdeI. The patent describes probes

CC (DNA or RNA) and their complements capable of detecting one or a

CC combination of HPV types 6, 11, 16, 18, 31, 33 and 35.

Sequence 591 BP; 148 A; 91 C; 121 G; 231 T;

Query Match 100.0%; Score 10; DB 1; Length 591;

Best Local Similarity 100.0%; Pred. No. 6e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10
 |||||
 DB 210 TGTGTCTCAG 219

RESULT 5

ID N92624 standard; CDNA; 1900 BP.

AC N92624;

DE 14-MAY-1990 (first entry)

DE Nucleotide sequence encoding mumps virus HN protein.

KW Haemagglutinin-neuraminidase; mumps; vaccine; ss.

OS Mumps virus.

FS Key

FT cds

Location/Qualifiers

91..1836

/tag= a

PN WO8911534-A.

PD 30-NOV-1989.

PF 17-MAY-1989; U02130.

PR 24-MAY-1988; US-198451.

PA (TEXA) Univ of Texas Syste.

PI Maxham MN;

WPI: 89-370727/50.

DR P-PSDB: P93666

PT HN specific degenerate oligonucleotide probe - useful for screening host

PT cells producing HN to use prod. in vaccine development

PS Claim 1; Page 25; 40pp; English.

CC Nucleotide sequence encoding HN protein of mumps virus, which may be used

CC to generate a large amounts of purified peptides from a unicellular host.

Sequence 1900 BP; 569 A; 433 C; 365 G; 533 T;

Query Match 100.0%; Score 10; DB 1; Length 1900;

Best Local Similarity 100.0%; Pred. No. 6.7e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTCTCAG 10
 |||||
 DB 1513 TGTGTCTCAG 1522

RESULT 6

ID N90647

ID N90647 standard; DNA; 5518 BP.
 AC N90647;
 DT 01-AUG-1990 (first entry)
 DE Nucleotide sequence of plasmid PBG396 including soluble T4-like (ST4)
 DE polypeptide number 12 (ST4#12)
 KM HIV; soluble T4-like polypeptide 12; immunotherapeutic; prophylactic;
 KW plasmid PBG396; diagnostic.
 OS Homo sapiens.
 PN W08901940-A.
 PD 09-MAR-1989.
 PF 01-SEP-1988; US-02940.
 PR 07-JAN-1988; US-141649, US-094322.
 PA (BIOJ) Biogen Inc.
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
 DR WPI; 89-085519/11.
 PT DNA sequences coding for soluble T4-like polypeptide(s) -
 PT used in immuno:therapeutic and immunosuppressive compns. and for
 PT preventing, treating or detecting AIDS
 PS Disclosure; ; 207pp; English.
 CC It is the nucleotide sequence of the plasmid PBG396. The sequence
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T
 CC cell tumour line REX and a lambda gt10 cDNA library derived from
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically
 CC synthesised antisense oligonucleotide DNA probes based on the known T4
 CC protein sequence was used. The polypeptide encoded is useful in
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used
 CC to purify HIV from a sample.
 SQ Sequence 5518 BP; 1334 A; 1424 C; 1403 G; 1357 T;

Query Match 100.0%; Score 10; DB 1; Length 5518;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
 |||||||||
 Db 26 TGTGTGTGAC 35

RESULT 7
 N90648
 ID N90648 standard; DNA; 5566 BP.
 AC N90648;
 DT 01-AUG-1990 (first entry)
 DE Nucleotide sequence of gene encoding mature T4 sequence of plasmid PBG393
 DE including soluble T4-like (ST4) polypeptide number 8 (ST4#8)
 KM HIV; soluble T4-like polypeptide 8; immunotherapeutic; prophylactic;
 KW plasmid PBG393; diagnostic.
 OS Homo sapiens.
 PN W08901940-A.
 PD 09-MAR-1989.
 PF 01-SEP-1988; US-02940.
 PR 07-JAN-1988; US-141649, US-094322.
 PA (BIOJ) Biogen Inc.
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
 DR WPI; 89-085519/11.
 PT DNA sequences coding for soluble T4-like polypeptide(s) -
 PT used in immuno:therapeutic and immunosuppressive compns. and for
 PT preventing, treating or detecting AIDS
 PS Disclosure; ; 207pp; English.
 CC It is the nucleotide sequence of region encoding the first 182 AA of the
 CC mature T4 sequence of plasmid PBG393. It does not contain the additional
 CC non-T4 6 amino acids at the C-terminus following AA 182. The sequence
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T
 CC cell tumour line REX and a lambda gt10 cDNA library derived from
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically
 CC synthesised antisense oligonucleotide DNA probes based on the known T4
 CC protein sequence was used. The polypeptide encoded is useful in
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used
 CC to purify HIV from a sample.
 SQ Sequence 5566 BP; 1350 A; 1437 C; 1413 G; 1365 T; 1 Others;

Query Match 100.0%; Score 10; DB 1; Length 5566;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
 |||||||||
 Db 26 TGTGTGTGAC 35

RESULT 8
 N90649
 ID N90649 standard; DNA; 5413 BP.
 AC N90649;
 DT 01-AUG-1990 (first entry)
 DE Nucleotide sequence of region encoding first 131 AA of plasmid PBG395
 DE including soluble T4-like (ST4) polypeptide number 10 (ST4#10)
 KM HIV; soluble T4-like polypeptide 10; immunotherapeutic; prophylactic.
 OS Homo sapiens.
 PN W08901940-A.
 PD 09-MAR-1989.
 PF 01-SEP-1988; US-02940.
 PR 07-JAN-1988; US-141649, US-094322.
 PA (BIOJ) Biogen Inc.
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
 DR WPI; 89-085519/11.
 PT DNA sequences coding for soluble T4-like polypeptide(s) -
 PT used in immuno:therapeutic and immunosuppressive compns. and for
 PT preventing, treating or detecting AIDS
 PS Disclosure; ; 207pp; English.
 CC It is the nucleotide sequence of the plasmid PBG395. The sequence
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T
 CC cell tumour line REX and a lambda gt10 cDNA library derived from
 CC peripheral blood lymphocytes (PBL). For screening a series of chemically
 CC synthesised antisense oligonucleotide DNA probes based on the known T4
 CC protein sequence was used. The polypeptide encoded is useful in
 CC immunotherapeutic, prophylactic and diagnostic compns. It can be used
 CC to purify HIV from a sample.
 SQ Sequence 5413 BP; 1309 A; 1401 C; 1365 G; 1338 T;

Query Match 100.0%; Score 10; DB 1; Length 5413;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
 |||||||||
 Db 26 TGTGTGTGAC 35

RESULT 9
 N90646
 ID N90646 standard; DNA; 5365 BP.
 AC N90646;
 DT 01-FEB-1991 (first entry)
 DE Nucleotide sequence of region encoding first 113 AA of plasmid PBG394
 DE including soluble T4-like (ST4) polypeptide number 9 (ST4#9)
 KM HIV; soluble T4-like polypeptide 9; immunotherapeutic; prophylactic;
 KW plasmid PBG394; diagnostic.
 OS Homo sapiens.
 PN W08901940-A.
 PD 09-MAR-1989.
 PF 01-SEP-1988; US-02940.
 PR 07-JAN-1988; US-141649.
 PA (BIOJ) Biogen Inc.
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR;
 DR WPI; 89-085519/11.
 PT DNA sequences coding for soluble T4-like polypeptide(s) -
 PT used in immuno:therapeutic and immunosuppressive compns. and for
 PT preventing, treating or detecting AIDS
 PS Fig 19; ; 207pp; English.
 CC It is the nucleotide sequence of the plasmid PBG394. The sequence
 CC was isolated from 2 libraries: a lambda gt cDNA library derived from T

CC cell tumour line REX and a lambda gt10 cDNA library derived from
 CC peripheral blood lymphocytes (PBL). For screening, a series of chemically
 CC synthesised antisense oligonucleotide DNA probes based on the known T4
 CC protein sequence was used. The polypeptide encoded is useful in
 CC immunotherapeutic, prophylactic and diagnostic compsns. It can be used
 CC to purify HIV from a sample.
 SQ Sequence 5365 BP; 1300 A; 1380 C; 1355 G; 1330 T;

Query Match 100.0%; Score 10; DB 1; Length 5365;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTCAG 10
 |||||
 DB 26 TGTGTGTCTCAG 35

RESULT 10
 590

004690 standard; cDNA; 1834 BP.

DE 05-OCT-1990 (first entry)
 DE Encodes Mammalian amino acid dehydrogenase activating factor-eta
 KM Mammalian amino acid dehydrogenase activating protein-eta;
 KM tyrosine dehydrogenase; tryptophan dehydrogenase; ss.
 OS synthetic.

FT Key Location/Qualifiers
 FT polyA_signal 1671..1676

FT cds /*tag= a
 166..1006
 /*tag= b

PN J02111796-A.
 PD 24-APR-1990.
 PF 21-OCT-1988; 264097.
 PR 21-OCT-1988; JP-264097.
 PA (TOFU) Toa Nenryo Kogyo KK.
 DR WPI: 90-169110/22.
 DR P-PSDB; R05084.

PT Amino acid dehydrogenase activating protein -
 PT isolated from mammal brain and activates tyrosine hydrogenase
 PS and tryptophan hydrogenase
 PS Disclosure: P; Japanese.

CC Probable error in specification at posn 168. C residue should read
 CC G (to give Met codon ATG instead of ATC).
 CC Encodes a protein characterised by Mr of 27KD (by SDS-PAGE), iso-
 CC electric point of 4.6 and an N-terminal amino acid sequence beginning
 MetGlyAsp.

CC Protein activates Tyr hydrogenase and Trp hydrogenase in the presence
 CC of calcium ions and calmodulin-dependent protein phosphatase Type II.
 CC Useful as a pharmaceutical analysis reagent.
 CC See also 004691 and 004692.

SQ Sequence 1834 BP; 546 A; 454 C; 471 G; 363 T;

Query Match 100.0%; Score 10; DB 1; Length 1834;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTCAG 10
 |||||
 DB 1589 TGTGTGTCTCAG 1598

RESULT 11

004555 standard; cDNA; 7316 BP.

DT 02-OCT-1990 (first entry)
 DE Plasmid p170.2 encoding T4 protein.
 CC T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;
 KM angiotensin; fusion protein; ds.
 FH Key Location/Qualifiers

FT signal_peptide 1199..1273
 FT /*tag= a
 FT mat_peptide 1274..2561
 FT /*tag= b
 FT /product=T4 protein
 FT replace (1457..1459,tgg)
 FT /*tag= c
 FT note="differs from M12807"
 FT replace (1958..1960,tty)
 FT /*tag= d
 FT note="differs from M12807"

PN W09004414-A.
 PD 03-MAY-1990.
 PF 18-OCT-1989; U04584.
 PR 18-OCT-1988; US-259355.
 PA (BIOJ) Biogen Inc.
 PI Meade HM, Lobb RR, Gates LL, Winkler G;
 DR WPI: 90-163876/21.
 DR P-PSDB; R04910.

PT New immunotoxin contg. soluble T4 protein components and toxin -
 PT esp. Pseudomonas endotoxin A, for treating or controlling AIDS
 PT and related conditions, and new DNA sequences.

PS Disclosure: pp: English.
 CC The T4 cDNA of p170.2 is almost identical to that reported by P.J.
 CC Maddon et al [Cell, 42, pp 93-104 (1985)]. The Maddon sequence was
 CC revised in 1988 to correct a DNA sequencing error at the codon for AA 3
 CC (corrected from Asp to Lys; see M12807 in Genbank). The DNA may be
 CC truncated (to remove transmembrane and intracellular regions) and/or
 CC modified by SDM, pref. so the prod. extends from AAs 3-183 of the
 CC mature protein. This DNA can then be ligated to a toxin DNA esp.
 CC angiotensin, or a fragment of pseudomonas exotoxin A contg. the
 CC translocation and ADP-ribosylation domains. The hybrid DNA can then be
 CC inserted into an expression vector and used to produce recombinant fusion
 CC protein which is useful for preventing or treating AIDS, ARC, and HIV
 CC infections. The T4 protein is an HIV receptor which binds to the virus
 CC or to infected cells carrying gp120/160 marker antigen, so provides v.
 CC specific targeting with minimal damage to non-target cells.

SQ Sequence 7316 BP; 1736 A; 1945 C; 1894 G; 1741 T;

Query Match 100.0%; Score 10; DB 1; Length 7316;
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGTGTCTCAG 10
 |||||
 DB 22 TGTGTGTCTCAG 31

RESULT 12

NB1632 standard; DNA; 828 BP.

DT 08-NOV-1990 (first entry)
 DE SV40 initial promoter region.
 KM Immunoglobulin H-chain enhancer; SV40; interferon; urokinase;
 KM hepatitis-B surface antigen; ss.
 PN J63160587-A.
 PD 04-JUL-1988.
 PF 23-DEC-1986; 305467.
 PR 23-DEC-1986; JP-305467.
 PA (GREC) Green Cross Corp.
 DR WPI: 88-224895/32.

PT Expression vector used for protein prodn. - comprises combination
 PT of immunoglobulin H-chain enhancer region and SV40 initial promoter
 PT region.
 PS Claim 3: page 2; 15pp; Japanese.

CC This sequence is inserted into an expression vector alongside the
 CC human immunoglobulin heavy chain enhancer gene and the construct is
 CC useful for the prodn. of proteins, eg interferon, urokinase or
 CC hepatitis-B. See also NB1631.
 SQ Sequence 828 BP; 220 A; 188 C; 191 G; 229 T;

Query Match 100.0%; Score 10; DB 1; Length 828;
 Best Local Similarity 100.0%; Pred. No. 6.2e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
 |||||
 Db 53 TGTGTGTGAC 62

RESULT 13
 NR80632
 ID NR80632 standard; DNA; 1606 BP.
 AC NR80632;
 DT 08-OCT-1990 (first entry)
 DE Human Bone Morphogenic Protein-2 class I cDNA.
 KW Bone morphogenic protein; hBMP-2 class I; probes: cartilage formation;
 bone formation; osteogenic cpds.; prodontal disease; ss.
 KW Homo sapiens.
 Key Location/Qualifiers
 cds 356..1543
 /*tag= a
 /product=hBMP-2 class I

MO8800205-A.
 PD 14-JAN-1988;
 PF 30-JUN-1987; U01537.
 PR 26-MAR-1987; US-031346,
 PA (GENE-) Genetics Inst Inc.
 PI Mozney JM, Rosen VA;
 DR WPI; 88-021565/03.
 DR P-PSDB; P80619.
 PT Bone morphogenic proteins - obtd. using recombinant DNA and used
 PT for inducing cartilage and bone formation.
 PS Disclosure: 7pp; English.
 CC The HindIII-SacI bovine genomic hBMP-2 fragment described in N80627
 CC is subcloned into M13, labelled and used as probe to screen
 CC polyadenylated RNAs from various cells and tissue sources.
 CC Sequence analysis of the strongly hybridising clones hBMP-2 class I
 CC (=BMP-2) indicated that they have extensive homology with the sequence
 CC given in N80622.
 CC The partial sequence is compiled from lambda U209-39 and several
 CC other hBMP-2 class I cDNA recombinants.
 CC This human cDNA hBMP-2 class I contains an open reading frame of 1188 bp,
 CC encoding a protein of 396 amino acids. The protein is preceded by a
 CC 5' untranslated region of 342 bp with stop codons in all frames. The 13
 CC bp region preceding this 5' untranslated region represents a linker used
 CC in the cDNA cloning procedures.
 See also NR80619-NR80636 and NR1963-64.
 Sequence 1606 BP; 399 A; 430 C; 423 G; 354 T;

Query Match 100.0%; Score 10; DB 1; Length 1606;
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGTGTGAC 10
 |||||
 Db 14 TGTGTGTGAC 23

RESULT 14
 NR1970/C
 ID NR1970 standard; DNA; 7533 BP.
 AC NR1970;
 DT 12-OCT-1990 (first entry)
 DE Plasmid pEMpl-CPA.
 KW tissue plasminogen activator; myocardial infarction; thrombosis;
 KW embolism; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_rna 7..1748
 FT /*tag= a
 FT /label=tpa_cDNA

FT cds 20..1708
 FT /*tag= b
 FT /product=t-PA
 FT 3'utr 1749..1955
 FT /*tag= c
 FT /label=R.3'UT
 FT /note="from Ig kappa light chain gene"
 FT 1930..1935
 FT poly-a-signal
 FT /*tag= d
 FT 1956..2104
 FT misc_rna
 FT /*tag= e
 FT /note="kappa 3 flanking sequence"
 FT 2105..4463
 FT /*tag= f
 FT /label-PBR322
 FT /note="contg. ampicillinase gene and origin of
 FT replication"
 FT 4464..6971
 FT misc_rna
 FT /*tag= g
 FT /label-PSV2-gpt
 FT /note="non-coding strand"
 FT 6437..6547
 FT protein_bind
 FT /*tag= h
 FT /label-cap_site
 FT /note="from SV40(gpt)"
 FT 6453..6469
 FT misc_rna
 FT /*tag= i
 FT /label-origin_region
 FT /note="from SV40"
 FT 6548..6619
 FT repeat_unit
 FT /*tag= j
 FT /rpt_type=direct
 FT 6620..6691
 FT repeat_unit
 FT /*tag= k
 FT /rpt_type=direct
 FT 6633..6650
 FT enhancer
 FT /*tag= l
 FT /note="from SV40"
 FT 6716..6945
 FT misc_rna
 FT /*tag= m
 FT /label-blocking_element
 FT /note="from lambda promoter"
 FT 6946..6971
 FT protein_bind
 FT /*tag= n
 FT /label-cap_site
 FT /note="lambda mRNA"
 FT 6972..7281
 FT enhancer
 FT /*tag= o
 FT /note="from Igh"
 FT 7282..7466
 FT promoter
 FT /*tag= p
 FT /note="from metallothioneine I"
 FT 7467..7533
 FT protein_bind
 FT /*tag= q
 FT /label-cap_site
 FT /note="from metallothioneine I"

MO8800242-A.
 PD 14-JAN-1988;
 PF 25-JUN-1987; U01569.
 PR 26-JUN-1986; US-879038.
 PA (DAWO-) Damon Biotech Inc.
 PI Gillies SD;
 DR WPI; 88-021592/03.
 DR P-PSDB; P81913.
 PT Prod. of human tissue plasminogen activator - by myeloma cells,
 PT pref. in presence of epsilon amino caproic acid.
 PS Disclosure: P; English.
 CC The plasmid was prepd by cleaving the very long 3' UT of t-PA cDNA
 CC (which causes mRNA instability) 34 nucleotides downstream of stop
 CC codon and inserting it into pEMpl vector. The pEMpl vector was
 CC constructed from the following components (see feature table):
 CC (a) a 2.25 PvuII-BamHI fragment from PSV2-gpt contg. the SV40
 CC enhancer and early region promoter, the E.coli gpt gene, the SV40

CC small tumour antigen intervening sequence, and the SV40 termination
 CC and polyadenylation signals;
 CC (b) a 2.3 kb PvuII-EcoRI fragment from pBR322 contg. the
 CC ampicillinase gene and the bacterial origin of replication;
 CC (c) a 0.3 kb PvuII-EcoRI fragment contg. an Ig heavy chain enhancer;
 CC (d) a 0.25 kb SacI-BglII fragment contg. the metallothionein I promoter;
 CC (e) a 0.4 kb AatII-HaeIII fragment from the 3' UT of Ig kappa light
 CC chain gene; and
 CC (f) a 0.26 kb XbaI-BstNI fragment contg. the Ig light chain promoter,
 CC TATAA sequence, and the initiation site.
 CC The expression plasmid can be used to transform host cells esp.
 CC J5581, ATCC CRL 9132.
 SO Sequence 7533 BP; 1918 A; 1892 C; 1847 G; 1875 T; 1 Others;

Query Match 100.0%; Score 10; DB 1; Length 7533;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGTGTGTGAC 10
 |||||
 6705 TGTGTGTGAC 6696

RESULT 15

ID N80979 standard; DNA; 600 BP.
 AC N80979;
 DT 12-SEP-1990 (first entry)
 DE Sequence of SV40 early promoter, polylinker and SV40 poly A addition
 DE region from pSV7d
 KM Platelet-derived growth factor (PDGF); SV40 virus; pSV7d;
 KM wound treatment; SV40 polylinker; SV40 poly A addition; ds.
 OS SV40 virus.
 FH Key
 FT Location/Qualifiers
 FT 1..315
 FT /tag= a
 FT /note="SV40 early promoter"
 FT 316..579
 FT /tag= b
 FT /note="SV40 polylinker"
 FT 580
 FT /tag= c
 FT /note="base 4210 of pBR322"
 FT 580..600
 FT /tag= d
 FT /note="pBR322 DNA"
 EP-288307-A.
 26-OCT-1988.
 PR 22-APR-1988; 303658.
 PA (CHIR-) Chiron Corp.
 PI Helfin C-H, Betscholtz C, Westermarck B, Knott TJ, Scott J, Bell GI,
 PI Rall L;
 DR MPI: 88-30127/43.
 PT Recombinant prodn. of growth factor A-chain polypeptide -
 PT for use in treating wounds and ulcers caused by diseases.
 PT Infections or neoplasms
 PS Example; Fig 7; 31pp; English.
 CC The mammalian cell shuttle vector plasmid pSV7d contains the SV40 origin
 CC of replication and early promoter (315 bp, PvuII pos 272-StuI pos 5193
 CC with an 8 bp deletion between nucleotides 173 and 182), a polylinker,
 CC and the SV40 poly A addition site (217 bp BclI pos 2775-pos 2558).
 CC The SV40 sequences were cloned into the pBR322 derivative pML between
 CC nucleotide 4210 and NruI pos 973. The SV40 sequences are positioned
 CC such that the direction of transcription from the early promoter is in
 CC the same direction as the ampicillin gene of the vector. Three separate
 CC platelet-derived growth factor mammalian cell expression vectors were
 CC constructed by isolating EcoRI fragments from cDNA clones and ligating
 CC them into pSV7d previously digested with EcoRI and treated with alkaline
 CC phosphatase. The resulting clones were called pSV7d-PDGF-A103 (D1),
 CC pSV7d-PDGF-A102 (13-1), and pSV7d-PDGF-B1 (B-chain). Recombinant PDGF is
 CC applied topically to wounds such as cutaneous, dermal, mucosal or

CC epithelial wounds. It may be used to treat any type of full or partial
 CC thickness wounds.
 SO Sequence 600 BP; 166 A; 158 C; 118 G; 158 T;

Query Match 100.0%; Score 10; DB 1; Length 600;
 Best Local Similarity 100.0%; Pred. No. 6e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGTGTGTGAC 10
 |||||
 Db 9 TGTGTGTGAC 18

Search completed: September 18, 1999, 00:33:31
 Job time: 18971 sec

Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillaceae;
 Bacillus.
 1 (bases 1 to 148068)
 Ogasawara, N.
 Direct Submission
 Submitted (03-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nucleotide
 Ogasawara, Nara Institute of Science and Technology, Graduate
 School of Biological Sciences, 8916-5 Takayama-cho, Ikoma, Nara
 630-01, Japan (E-mail:nogasawa@bs.aist-nara.ac.jp,
 Tel:07437-2-5430, Fax:07437-2-5439)
 2 (sites)
 Kasahara, Y., Nakai, S., Lee, S., Sadate, Y. and Ogasawara, N.
 A 148 kbp sequence of the region between 35 and 47 degree of the
 Bacillus subtilis genome
 Unpublished (1997)
 JOURNAL
 3 (sites)
 Bejot, C., Ayora, S., Exley, R., Hirschbein, L., Ogasawara, N.,
 Kasahara, Y., Alonso, J. C. and Hegarat, F. L.
 Characterization of an *irp*-like (*irpC*) gene from *Bacillus subtilis*
 Mol. Gen. Genet. 236 (1), 63-71 (1997)
 98000887
 source
 Location/Qualifiers
 1. 148068
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 /map="35 - 47 degree"
 1. 33
 /note="PROBABLE RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 175. 2273
 /gene="ydaA"
 175. 180
 /gene="ydaA"
 189. 2273
 /gene="ydaA"
 /note="PROBABLE HTH_LYSR_FAMILY TRANSCRIPTIONAL
 REGULATOR."
 /codon_start=1
 /transl_table=11
 /protein_id="BA19254.1"
 /db_xref="PID:d1020007"
 /db_xref="PID:g1881227"
 /db_xref="GI:1881227"
 /translation="MYMTAEORLKLKLLQNRIVTLELAEIMOVSTRIRHLSKI
 KPMETVGLTLDKOPGKLAAGSPGKQKLLDLSGEHEYSADERKLLILCSLES
 QEPVKTTLAHLDQVNAIVSYDDELEKWIISFGLLIRKRGQIOLIGPENAKRKI
 VGNLIYRDLIOMLEAVEINIKGTDSSEKMGVYSGKELLMERILPOLAKIKAF
 LSDSYALVHLTYATERKIGETITMENONELELNNAKEIYSSADELAELEBRAGV
 TIPBAEYGITIHRSANRKYKTEYKQELTELETAQTKRLIAFISDKIRMDLTKNYS
 LYBGLNHEPFAVRKIKENETINPKMEQIKROYFLYNAIEGVKKEIPGMSFSDE
 IAFILVLFHSGALEIKKEAKVKAIVVSSGSKMLASRLKKELEIEISFSSSIE
 LKGDQVADMIIVSTVPIPYENIDIVMSPLNEEDANVOKYIKRIRPIILKKRSS
 KEEOQADVDMLAEASISGRVMEVIDVLRHFLIOLKTPNDHSMILLLELPOLKQD
 GLIRDPKAAVCLAEKOGSLGIPGNALYLKNDIYLPFFKMPDSTPIEVGDM
 DGLTRKTRILVMAPOSLSAESGEIISAISSATISGESMAQFOEGGEIYORLNR
 IFTWKEKNIL"
 2472. 3881
 /gene="ydaB"
 2472. 2479
 /gene="ydaB"
 2484. 3881
 /gene="ydaB"
 /note="SIMILAR TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 COVALENT BINDING OF AMP TO THEIR SUBSTRATE."
 /codon_start=1
 /transl_table=11
 /protein_id="BA19255.1"
 /db_xref="PID:d1020008"
 /db_xref="PID:g1881228"
 /db_xref="GI:1881228"
 /translation="METIOHLLHDMNPSEIEPAVKSQDHTLTLYKGYRRINOLANAM
 LKGIQKGRVALCKNGHPASTVMPAALEIGAVVVPVSNQKPYEMTIGILKASERKA
 MYEAEKELIDVLELSSLCVTMETGAYETSAFEALFACPDHLPETEMVSPDPT
 ALMFTSGITGNKRCMTTHGIRYVKNSSIAARKGIRFLACHPITHTSLICIM
 LGTAEETFEVFKDQDPVHMLKYLEEKIOTVALPVTYLLLEAMKHOTDSLVY
 LMTGKTRPSSLSIRYLDIGIPLAHGSGTEAGISWPPDMMDKRAASKGVACAK
 VKVEDLPLGKELPOGEGEIVYTPPLFKYEENPEPATAVLONGFRITDSSVYED
 GFPTTIRKIDVITVIGDNNYTPQVEEVIQITGILETAVAGIPDPYGERKPAFTVK
 NGQRTIEEDVIAFL"
 complement(4016. 4577)
 /gene="ydaC"
 complement(4016. 4561)
 /gene="ydaC"
 /note="FUNCTION UNKNOWN, SIMILAR PRODUCT IN SEVERAL
 BACTERIA."
 /codon_start=1
 /transl_table=11
 /protein_id="BA19256.1"
 /db_xref="PID:d1020009"
 /db_xref="PID:g1881229"
 /db_xref="GI:1881229"
 /translation="MAGYIMAAENOTLNQWITNOLGTRPDSILEVGFPGYCMQOM
 IKREKDVHLHGIDVSEAMLKLAARRVYKPGVRLIGSIEFPPLPAFEDKVISVNYT
 IWDQTKGKQIRALKRPGSKAITQPPADADSPKRTKSGFQMLADRAAFEDID
 IORNKIRPELSVCATKKPAT"
 complement(4572. 4577)
 /gene="ydaC"
 4752. 5629
 /gene="ydaC"
 4752. 5629
 /gene="ydaC"
 4752. 4759
 /gene="ydaA"
 4769. 5629
 /gene="ydaA"
 /note="BELONGS TO THE INSECT-TYPE ALCOHOL DEHYDROGENASE /
 RIBITOL DEHYDROGENASE FAMILY."
 /codon_start=1
 /transl_table=11
 /protein_id="BA19257.1"
 /db_xref="PID:d1020010"
 /db_xref="PID:g1881230"
 /db_xref="GI:1881230"
 /translation="MANYPKELPAOTROSROPGIESPMNPVEYEDYKADKLRV
 ALITGDSIGRAVSAVAKEGADIAIVYKDEHDEEETKRRYQEGCVCLTAGVG
 BEEFCNEAVKTVKELGGLDILVNNAGEQHPESINDITSEDLHRFKTFYSQFLT
 KKAIDLKPSAINTSINPYVGNPTLIDYATKCAINAFITMAQALVKDRIYNA
 VAPGPWIPILIPATPEETVAFQGDTPMGRPGQPYEHGCVYLASDESSVYTGTL
 HNGSGNFVT"
 5628. 6148
 /gene="ydaB"
 5628. 3639
 /gene="ydaB"
 5645. 6148
 /gene="ydaB"
 /note="FUNCTION UNKNOWN."
 /codon_start=1
 /transl_table=11
 /protein_id="BA19258.1"
 /db_xref="PID:d1020011"
 /db_xref="PID:g1881231"
 /db_xref="GI:1881231"
 /translation="MGITKEVNSYQKAGIYLTDEVDQIOLMDYGLCKERRVGIOQL
 ETVVNDRIKSELVLFPGQTCPEHRHPVDQDEKQCFRCRYGKGVLYLYGGEKTPPL
 PVLPPQEDREHYTWHEILEPFGQYITPNTKHFQAGEGAVVTEMSYSTDKHD
 IFTDPRI"
 6146. 6174
 /note="PROBABLE RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR."
 6219. 6784
 /gene="ydaF"
 6219. 6230
 /gene="ydaF"
 6233. 6784
 /gene="ydaF"
 /note="PROBABLE ACETYLTRANSFERASE."
 /codon_start=1

JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	FEATURES source	gene CDS	gene CDS	gene CDS
FEBBS Lett. 363 (1-2), 75-77 (1995)	FEBBS Lett. 356 (2-3), 330-332 (1994)	Hayashi, M., Hirai, R. and Unemoto, T. Cloning of the Na(+)-translocating NADH-quinone reductase gene from the marine bacterium <i>Vibrio alginolyticus</i> and the expression of the beta-subunit in <i>Escherichia coli</i>	Hayashi, M., Nakayama, Y. and Unemoto, T. Existence of Na+-translocating NADH-quinone reductase in <i>Haemophilus influenzae</i>	Nakayama, Y., Hayashi, M. and Unemoto, T. Identification of six subunits constituting Na+-translocating NADH-quinone reductase from the marine <i>Vibrio alginolyticus</i>	On Oct 22, 1997 this sequence version replaced gi:677955 gi:803412. D43958: submitted (17-Dec-1995) D49364: submitted (17-Feb-1995).	1. 6908 /organism="Vibrio alginolyticus" /db_xref="taxon:663" /note="There are conflicts between this sequence and GenBank Accession Number:237111." 386. 1726 /gene="nqr1" 386. 1726 /gene="nqr1" /note="contains FMN as a cofactor; nqrA, Nqr1 subunit of Na-NQR complex alpha-subunit" /codon_start=1 /transl_table=11 /evidence=experimental /product="Nqr1 subunit of Na-translocating NADH-quinone reductase complex alpha-subunit" /protein_id="BAa22910.1" /db_xref="PID:d1023781" /db_xref="PID:g2558473" /db_xref="GI:2558473" /translation="MTTKKGLDLPFAGPPSOYINDNGKITKKYALLGEETVGRKPLPMHVRGDEVKRAQVLEEDKNRQVAFETAPAGKATVEVNRGAKRVLAQSVLIEVAGEQVTFDKEALQSLGDEEVIRKQLVDSGLWTALTRFSPVPIESTKAIPTYAMDNPITAKRELLINQOEAFVIGLIDILSLATEGKYVCGSLPSSQSNVEEHFDPDHPAGLAEIHNFILYPAENVAENSVINQDVIARGLKLEINGETIDPRVSIAGVYNNRPLRVTVGAGLIDDLTDLNEMLNPGEVRIYSGVLTGRTAGTGHATIGRIHQOVSIVRESREKELFGHARGKNTFSYTRKFLGHVERGQLENNNTTNGSDRSMPVIGNETVRVMDMEPTLLNSDLADGDTDSQAALGALELDEDLALCTFCQPKRYEYGLLEKLCIDTIEREG"	1732. 2976 /gene="nqr2" 1732. 2976 /gene="nqr2" /note="hydrophobic membrane protein with pI 8.14; nqrB, Nqr2 subunit of Na-NQR complex" /codon_start=1 /transl_table=11 /evidence=experimental /product="Nqr2 subunit of Na-translocating NADH-quinone reductase complex" /protein_id="BAa22911.1" /db_xref="PID:d1023782" /db_xref="PID:g2558474" /db_xref="GI:2558474" /translation="NALKKRFLEDEIHPEPGKHEKRALYENAVATVEYTPGIVTNKSSHRDSDVDKRLIMINWEVAPFMPEGMNNAAGQAIALANHYADQQLATVVISGNMXYWTEEMGQITADAGVGSKMLGATFPLIYATVPLVGGFWELFCMWRKHEVNGEFTVISLLEALIPPTLLPLMQAALGITFGVYVAKIEGQTGNFLNPLAGARAFLEFAFVATISGALVYVNRADGEGSATLISOMAGGNAALVNTVGTGSPITWMDAFIGTIPGSGIVS		

TLALMGAMVYRIASRIAGVMTGMAVSTLENVIGSDINPFMFHMLVIG
GFAFGFMFMAIDPVASFTNKGKMWYILIGAMCMVIRVNPAYPEGMALAILFANLE
APLDHVIIEKRIKRLARYGK"
/db_xref="PID:d1023786"
/db_xref="PID:g2558478"
/db_xref="GI:2558478"
/gene="ngt3"
2966. 3736
/gene="ngt3"
/note="acidic protein with pI 4.83; ngf, ngf3 subunit of
Na-NQR complex gamma-subunit"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="Ngt3 subunit of Na-translocating NADH-quinone
reductase complex gamma-subunit"
/protein_id="BAA22912.1"
/db_xref="PID:d1023783"
/db_xref="PID:g2558475"
/db_xref="GI:2558475"
/translation="MASNDSIKTKLGLVIGLSVCSIIYSTAAVGIKDKANAVD
KQSKIVEAGIDANGKVPLEFAVIEPRVLDLETGFTGNSASTYDORAKDAENS
IALPDEVDADIRIRANAVVYLVKDDDEVKILLPHGGLMSMAYFAVEIDGNT
VSAITTYEAGEIPGLGEVENSMDQFEGIKLYNEDHQAIAIKVKGADQSGSHGD
GLSGATLTSNCGHTFDFWLGDKGFPFLAKVRGEIN"
3736. 4368
/gene="ngt4"
3736. 4368
/gene="ngt4"
/note="hydrophobic membrane protein with pI 9.17; ngfD,
Ngt4 subunit of Na-NQR complex"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="Ngt4 subunit of Na-translocating NADH-quinone
reductase complex"
/protein_id="BAA22913.1"
/db_xref="PID:d1023784"
/db_xref="PID:g2558476"
/db_xref="GI:2558476"
/translation="MSSQNVKSKILAPVDLNNPILAVGVCSALAVTKLEAFVY
TLAVFTVLSNFSVLSIRHNPVSRIIYOMALIASIVYDVLKAYIDSKOLS
VFGLITINCIVMGRAEAFKAPSPSLIDGNGSGYGVITVGFPEELSGSL
FGLVPLVSGNGVTPGMLLAPSAFLIGLIVYIRIKPEQVAK"
4376. 4972
/gene="ngt5"
4376. 4972
/gene="ngt5"
/note="hydrophobic membrane protein with pI 7.00; ngfE,
Ngt5 subunit of Na-NQR complex"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="Ngt5 subunit of Na-translocating NADH-quinone
reductase complex"
/protein_id="BAA22914.1"
/db_xref="PID:d1023785"
/db_xref="PID:g2558477"
/db_xref="GI:2558477"
/translation="MEHYISLVKSIETENALSFIFGCTFLAVSKRYTSGSLVA
VVVLTAVPNVNIYNDKRNALVGVDSIFNFTIFGIVIALVOLLIEVDREF
PPLYNALGIFLPLITVNCALFGVSMVQDYNFASIVYFGSGVGMALAIYALAGI
REKMKSDVPEGLRGILGITFTGLMALGFMSFSGVQL"
5006. 6229
/gene="ngt6"
5006. 6229
/gene="ngt6"
/note="contains FAD as a cofactor; NADH dehydrogenase
activity; pI 4.55; ngfF, Ngt6 subunit of Na-NQR complex
beta-subunit"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="Ngt6 subunit of Na-translocating NADH-quinone
reductase complex beta-subunit"

/protein_id="BAA22915.1"
/db_xref="PID:d1023786"
/db_xref="PID:g2558478"
/db_xref="GI:2558478"
/translation="MDIIGVWFTLIVALLVILVFAKSKIVPTGDTITISVNDPSL
AATPGGKILSLAAGVYVSACGGCCGCRKRVKSGGDIIPEDHTTKEA
RSEERLACQVAMKTDMDIPEPITFVKKECTVINDNKATIKELKQIDGESVP
FRAGYIQIEAPAHVADYDIPEDYREDWENFNFRESKYNEETIRAYSNANPE
EHGIIINLVRIATPPNPDPVPPGSIYSYLSKEDDKTISGPEEFPAKTDAAEM
FVGGAGAPMRSHIFDOLKRLSKRMSFVGRSRKREMYVEDMDIAENDNFVW
HCAISDPLPEDNDYGTGFHNLVYENYLRDHAPECDCEYMGCPMAAAYIGMLKD
ICVEDNIIIDDDG"
6252. 6276
terminator
BASE COUNT 1840 a 1473 c 1619 g 1976 t
ORIGIN
Query Match 100.0%; Score 11; DB 1; Length 6908;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CTGTTCTAGT 11
|||||
Db 3856 CTGTTCTAGT 3846
RESULT 3
AB008698
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
JOURNAL
TITLE
AUTHORS
DIRECT SUBMISSION
SUBMITTED (10-NOV-1997) to the DDBJ/EMBL/Genbank databases. Satoh
Yamamoto, Marine Biotechnology Institute, Kamaiishi Laboratories;
Heita 3-75-1, Kamaiishi City, Iwate 026, Japan
(E-mail: syamamoto@kamaiishi.mbio.co.jp, Tel: 81-193-26-6544,
Fax: 81-193-26-6584)
FEATURES
source
location/Qualifiers
1. 1167
/organism="Acinetobacter radiorestrictus"
/strain="SEIP 12.81"
/db_xref="taxon:40216"
1. 1167
/gene="gyrB"
1. 1167
/gene="gyrB"
/RC_number="5.99.1.3"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B"
/protein_id="BAA75415.1"
/db_xref="PID:d1039153"
/db_xref="PID:g4514433"
/db_xref="GI:4514433"
/translation="DNSYVSGGLHGVSVVYNAALSSKILEIVORAGQIHEDEYOGV
POYPLRVGOTERTGTKRVFWSAEFTSOTIFNVDILARLRSLANGLVGLDCE
RINLHEVFEGLSEFVYINEGKNHNLNIEFTTQANGISGEVVALQWMSYQENV

RCFTNNIPKDGSTHLAGRAALTGRNLNENILKKKAVATGDDAREGATLAVS
VKYDPEKSSOTKREKTVSSSEVRAVEONAKKASELLEPAAKSIAGKIIDANAR
DARRKAREMTRRSADIDIALPKKLDCCOKDPALSELFLVEDSDSGSKOGRNRKM
QALPLKGLKILNERNRFRDPMISSAEVGLITLALGGGIGREYVNPDKLRHKIV"

BASE COUNT 317 a 230 c 302 g 318 t

Query Match 100.0%; Score 11; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11
|||||
Db 940 CTGTTCTAGT 950

RESULT 4
JUS AACG/c
DEFINITION ACCAA61G 1154 bp DNA BCT 17-FEB-1994
Acinetobacter haemolyticus aminoglycoside 6'-N-acetyltransferase
(aac(6')-I_g) gene, complete cds.
ACCESSION L09246
NID 9455436
VERSION L09246.1 GI:455436
KEYWORDS aac(6')-I_g gene; aminoglycoside 6'-N-acetyltransferase.
SOURCE Acinetobacter haemolyticus (strain BM2685) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
Moraxellaceae; Acinetobacter.

REFERENCE 1 (bases 1 to 1154)
AUTHORS Lambert, T., Gerbaud, G., Gallmand, M. and Courvaill, P.
TITLE Characterization of the Acinetobacter haemolyticus aac(6')-I_g gene
encoding an aminoglycoside 6'-N-acetyltransferase which modifies
amikacin

JOURNAL Antimicrob. Agents Chemother. 37, 2093-2100 (1993)
MEDLINE 94079352
FEATURES
Location/Qualifiers
1..1154
/organism="Acinetobacter haemolyticus"
/strain="BM2685"
/db_xref="taxon:29430"
544..981
/gene="aac(6')-I_g"
544..981
/gene="aac(6')-I_g"
/codon_start=1
/transl_table=11
/product="aminoglycoside 6'-N-acetyltransferase"
/protein_id="AA21889.1"
/db_xref="PID:9455437"
/db_xref="GI:455437"
/translation="MNKIPASASLKDLELRNKMSPSEASHLOEMHOLAETAYLO
LLAYSDHQAIAIMLASIRPEYVNGTSPVGLBEITVLPARRSGVATMLIRQAEVW
AKQPSCTEFASDAIDNVTISAMHRSLSGFETEKVYVSKID"

BASE COUNT 322 a 200 c 372 t
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 1154;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11
|||||
Db 432 CTGTTCTAGT 422

RESULT 5
ACCGYB12 300 bp DNA BCT 04-FEB-1999
LOCUS Acinetobacter sp. gene for DNA gyrase subunit B protein, C terminal
side, partial cds.

ACCESSION D73423
NID 913422104
VERSION D73423.1 GI:1322104
KEYWORDS DNA gyrase subunit B protein.
SOURCE Acinetobacter sp. (strain:SEIP 12.81) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
Moraxellaceae; Acinetobacter.

REFERENCE 1 (bases 1 to 300)
AUTHORS Yamamoto, S.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1995) to the DDBJ/EMBL/GenBank databases. Satoshi
Yamamoto, Marine Biotechnology Institute, Kamaiishi Laboratories;
3-75-1 Heita, Kamaiishi, Iwate 026, Japan
(E-mail:HGSD02251@infocenter.or.jp, Tel:0193-26-6538,
Fax:0193-26-6584)

REFERENCE 2 (bases 1 to 300)
AUTHORS Yamamoto, S.
JOURNAL Unpublished (1996)
REFERENCE 3 (sites)
AUTHORS Yamamoto, S. and Harayama, S.
TITLE Phylogenetic analysis of Acinetobacter strains based on the
nucleotide sequences of gyrB genes and on the amino acid sequences
of their products
Int. J. Syst. Bacteriol. 46 (2), 506-511 (1996)
97088999

JOURNAL Location/Qualifiers
MEDLINE 1..300
FEATURES
/organism="Acinetobacter sp."
/strain="SEIP 12.81"
/db_xref="taxon:472"
/note="correspond to pos.1183-1482 in E.coli K-12 gyrB
gene (X04341, X00870)."
1..300
/gene="gyrB"
1..300
/gene="gyrB"
/transl_table=11
/product="DNA gyrase subunit B protein"
/protein_id="BA11148.1"
/db_xref="PID:d1011809"
/db_xref="PID:91322105"
/db_xref="GI:1322105"
/translation="KSAIDINGLPKRIADQCKDPALSELFLVEGDSAGSANOGRNR
KMAIDLEPKGLNVERAFRPMISSAEVGLITLALGGGIGREYVNPDKLRHKIV"

BASE COUNT 75 a 57 c 86 g 82 t
ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11
|||||
Db 76 CTGTTCTAGT 86

RESULT 6
AF027868 87500 bp DNA BCT 15-NOV-1997
LOCUS Bacillus subtilis chromosome region between terc and odnab.
DEFINITION AF027868
ACCESSION AF027868
NID 92618993
VERSION AF027868.1 GI:2618993
KEYWORDS Bacillus subtilis.
SOURCE Bacillus subtilis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus.
REFERENCE 1 (bases 59208 to 59540)
AUTHORS Muller, J., Walter, F., van Dijk, J.M. and Behnke, D.

TITLE
Suppression of the growth and export defects of an *Escherichia coli* secA(Ts) mutant by a gene cloned from *Bacillus subtilis*

JOURNAL
Mol. Gen. Genet. 235 (1), 89-96 (1992)

REFERENCE
2 (bases 34592 to 35233)

AUTHORS
Wolf, M., Geczi, A., Simon, O. and Boriss, R.

TITLE
Genes encoding xylan and beta-glucan hydrolyzing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase

JOURNAL
Microbiology 141 (Pt 2), 281-290 (1995)

MEDLINE
95219081

REFERENCE
3 (bases 1 to 87500)

AUTHORS
Lapilus, A., Galleron, N., Sorokin, A. and Ehrlich, D.

TITLE
Sequence analysis of the *Bacillus subtilis* chromosome region between the *terc* and *odhAB* loci cloned in a yeast artificial chromosome

JOURNAL
Unpublished

REFERENCE
4 (bases 1 to 87500)

AUTHORS
Lapilus, A., Galleron, N., Sorokin, A. and Ehrlich, D.

TITLE
Direct Submision

JOURNAL
Submitted (01-OCT-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France

FEATURES
source
1. 87500
/location/Qualifiers
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
/gene="yoxB"
/gene="yoxB"
/citation={3}
/gene="yoxB"
/citation={3}
/codon_start=1
/transl_table=11
/product="yoxB"
/protein_id="AAB84418.1"
/db_xref="PID:g2618994"
/db_xref="GI:2618994"
/translation="AYETRAESLLOGGTLLEONAVDTKIAAGSGSMRNLITMDLGG
LPQAFQISAAALFGLRMLHTTSVIRPVPIQOKNYGXPAITHSPISTVAFGH
PPAELNHYVGPALLIPLGAAETVCELTIRHSLONLCLNCFPPHSGANGHRLKQFS
KNKEQLEROIRFEPALVDPAIEKRK"
604. 1212
/gene="yoxA"
/gene="yoxA"
604. 1101
/note="similar to B. subtilis P20 protein and E. coli
ribosomal protein alanine acetyltransferase (194 aa)"
/citation={3}
/codon_start=1
/transl_table=11
/product="putative alanine acetyl transferase"
/protein_id="AAB84419.1"
/db_xref="PID:g2618995"
/db_xref="GI:2618995"
/translation="MEPILEDRLRLROITDODARAFACFSNDEVTYVYGLNENESI
EQAISMIOTFALYQEKRGIRKGTIRKOTKLTIGTGFHALAQHRRREIEIETIPEH
WRNGFASEYISKVSYSFGSALGSLRIGAVETIDNASNRLIKMGFOREGLROYMS
KWDSI"
1187. 1212
/gene="yoxA"
complement(1217. 2461)
/gene="yoxB"
complement(1217. 2461)
/gene="yoxB"
/note="similar to B. subtilis multidrug resistance protein
2 (400 aa)"
/citation={3}
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAB84444.1"
/db_xref="PID:g2619020"

/db_xref="GI:2619020"
/translation="MLDKTIGPKRLANGLFVLPFMKDGLEQGLSPFLIENGLTVO
OSASITGTIGLALIASWFSVCLEARAKRTMTGLTFYIGTAPLTFVFEEDINLP
VMVYVYFVKGLGYPFLFALSYFLTWIYRTPOSRLSTAVGEMWIAVCLGFWGAMWSSY
AIKAFYGLNTLMSSIFWVCLGAFALFINKDRPEKRRKRSRTAEILKGVYITLFTNP
RVLGGIRIRINSIGTYGFPVFLPMHMAHGISNVMVLIQINGITFGLIVNLYFGYI
GDKRGKMTVIFWFGGCGITVLLYVAPVSGGSLAVSVIGTIFGGLIAGVPIGA
IVPVAGDKRAANSVNLNLAAGLSAFVGPALAMFLIGLVAGQVWIFRALLYASAVL
TKCHLPEKAKVKEISPOVNS"
complement(2555. 4018)
/gene="yoxC"
/gene="yoxC"
/citation={3}
complement(2555. 4018)
/gene="yoxC"
/note="similar to Lactobacillus xylulose kinase (501 aa)"
/citation={3}
/codon_start=1
/transl_table=11
/product="xylulose kinase"
/protein_id="AAB84445.1"
/db_xref="PID:g2619021"
/db_xref="GI:2619021"
/translation="MKKQGYLVEFDIGTGNARVAVSVTSGSVLTVREDEIYETETLY
PDSYFSPVOVAMKOVNMLAKRALSODIDITGLSTSORGIVLIDONSPFGLIPN
IDNNGRENAEIPMEETISYSGRLPRLAEALILYGLKORQPSLWERTASTISDM
VTYOLSGILYEPQATEETLFDVKNQNTSEKMDIFGFSILPLVRAQTAIGTIA
NEVASLGLSINAKVIAGGDTOLAVSTGSLDIVVSCGTPITITIDHGDTKA
KAMLNCHTDGHLVETNPGLTGNOKLOIFYPNETYEVAMEEISALAKEDNACVA
ALGSYLSERKALTRGGFLPDAPLPAHLKRAHFAVRLAEELAFISIKWEDITLTPF
ERDVMYCGGFGSKALQYIADLQKRYVQDEYRHOASVGAIVICNETQTLEENS
ANVAVIEPKDQIELEAEEMKQTFRFGSGESKVL"
complement(4036. 5070)
/gene="yoxD"
/gene="yoxD"
complement(4036. 5070)
/gene="yoxD"
/note="similar to B. subtilis D-3-phosphoglycerate
dehydrogenase (419 aa) encoded by *serA* gene"
/citation={3}
/codon_start=1
/transl_table=11
/product="yoxD"
/protein_id="AAB84446.1"
/db_xref="PID:g2619022"
/db_xref="GI:2619022"
/translation="MKNTMKRMFCMTVLVYAPYNEGRKLENLIFGSVAYQSKREG
RAYREDELIOLKATNATGLTLEDQVDSVFASVPELSEFGVCRGMPNSNDVAAASK
RGIVFYTPGNNAQAVAMETIGNISLPHRSASNOMKQEMSDVLOAVKRGNE
LGTQVMICMGACGRIAKLTITPAFDCKIKKYDPYIODDHLIYKASIKTYFSSDIY
SVHLPRTEITGLIDROYFDLMEKESALFVNTSRAVVNREDLTVLKEHKISGAILDV
FYHEPESDEELISLPNVLATPPLAGATEVEEDHVTILNKALKKMGKEXTLNIOTM
YNKDALKTGG"
5394. 7427
/gene="yoxE"
/gene="yoxE"
5394. 7427
/note="yoxE"
/note="similar to B. subtilis yoxE protein (667 aa) and
E. coli formate dehydrogenase (715 aa)"
/citation={3}
/codon_start=1
/transl_table=11
/product="putative formate dehydrogenase"
/protein_id="AAB84420.1"
/db_xref="PID:g2618996"
/db_xref="GI:2618996"
/translation="MRSFATQONGIFKVSCLDPCDQGLLIHKKDGKIYVQGDPDH
PYTAGNFCNKVNTMTERIYDEKRTITPKRTGAKGQALPEPISKEKALIDITSMKOL
IDEGASILDPYFYGNAKGLTARAGMRPFYRSGSQQLERTYCSKAGSGCYRTWGI
SAGIDPEIYHITKLFITWGINAVSTNHQITLQAKRKAKTIVIVDHNQGTGLND
WFIPKPGTDSALALGIMHILFEKNLHDEAFLEBYTYGELREHVAQYDPEKVSITIT
GVSTEDYIRLAKYGETSPSIFIRIGNPOHNDGNGMIVRTIACPATIGOMLHAGGA
IKHNSGILEYNTNALRQDPLKGRTPSPFNNOGQVRLTDPTRISLFTYGNPVAV
APEAKNRQGLREDLFTVVDLFLTETAAVADIVLPATSAFENTDYTSYWHYVYIOL
QOPIERYGSEKSTEVFRLLAEMGTDELDKSDDEVILRQALDHPDNHLEAIDID

SLTKSEMKAKREKPLFPGELPTPSGKILEYSEKMKODGFPALPTYPTVNDNEHPFM
YVGPNNHNFINSFNNEKHLEKTPKIFITKDAEKGIYDGAIPVRWNRGSECEL
TAAVEQVLPGVVVSOGMLADROGKTKACKRANPSLVNRGRRLFFRPCTD"

gene
terminator
complement(9379..11027)
/gene="yohA"
complement(9379..9406)
/gene="yohA"
complement(9423..11027)
/gene="yohA"
/note="similar to B. subtilis methyl-accepting chemotaxis protein McpA (661 aa)"
/citation=[3]
/codon_start=1
/transl_table=11
/product="yohA"
/protein_id="AAB84447.1"
/db_xref="pid:g2619023"
/db_xref="gi:2619023"
/translation="MISSVKKNEELKDKMEROKETKHITOVETLGLSNDREGLITRD
KEYDGMKAKADVLEKSDRVNDLIDEEYQSNIEDIKTSFYQYRALNDQVYTAISSN
PKRAETIHGEERTLRKEGVAPAVNKLSDRLQVEVDLDEIQNGKMSQLITVG
ISVILGIYLSIMILKSIMVPLKSIINKOLEIHAEDLTKVYINRKRDEFGQLAOSFN
SETHSLQIVKQISSSEQVAASEELASAESEKSTSEHISRAMOMADSNVKSMS
TEKSAESTRELDSTISSVANTGNITADSSMRDRAEIGSKVNMKLDQMKRIDKSDV
SAGNLIQTVASTAETSDISSLTITISEQPNILANAEARAGEOGCFVAVAEV
RKLADEINKSANHIOSVARTIONESIETVYNTIKYVOENVSCTIVASQETGNNFELN
LVEQVTSQIQEVAATQQLTSGEVNIQHTVHTLAGTKETSNANTEAVANSOEOHSM
GEISYAAESLSQIAELQTVINRFKY"
11755..12910
/gene="yohA"
/citation=[3]
11755..12882
/gene="yohA"
/note="similar to H. influenzae HI0035 protein (551 aa)"
/citation=[3]
/codon_start=1
/transl_table=11
/product="transcription antiterminalator"
/protein_id="AAB84421.1"
/db_xref="pid:g2618997"
/db_xref="gi:2618997"
/translation="MNTVMVSFASSAELKDKENCPPEHILDMYEOAKHDSFTHTF"

Query Match 100.0%; Score 11; DB 1; Length 87500;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTTCTACT 11
|||||
Db 65049 CTGTTCTACT 65039

RESULT 7
LOCUS BACS18422/c 2985 bp DNA BCT 23-NOV-1998
DEFINITION Bacillus sporothermodurans partial uvrB and uvrA genes.
ACCESSION Y184422
NID 93924674
VERSION Y184422.1 GI:3924674
KEYWORDS uvra gene; uvrb gene.
SOURCE Bacillus sporothermodurans.
ORGANISM Bacillus sporothermodurans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
1 (bases 1 to 2985)
REFERENCE
AUTHORS Herman, L. and Heyndrickx, M.
TITLE Bacterial typing by REP-PCR can be based on the presence of random
homologous sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2985)
AUTHORS Herman, L.M.F.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) L.M.F. Herman, Centre of Agricultural

Research, Department Animal Product Quality, Brusselssesteenweg 370,
9090 Melle, BELGIUM

FEATURES
source
Location/Qualifiers
1..2985
/organism="Bacillus sporothermodurans"
/strain="MB342"
/db_xref="taxon:46224"
<1..1784
/gene="uvrB"
/product="uvrB protein"
1..1784
/gene="uvrB"
1792..>2985
/gene="uvrA"
/product="uvrA protein"
1792..2985
/gene="uvrA"

gene
BASE COUNT 994 a 497 c 655 g 831 t 8 others

ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 2985;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTTCTACT 11
|||||
Db 2467 CTGTTCTACT 2457

RESULT 8
LOCUS BAPRPEG 4753 bp DNA BCT 24-JUN-1994
DEFINITION B. aphidicola trpE and trpG genes for anthranilate synthase,
components I and II.
ACCESSION Z21938
NID 5472880
VERSION Z21938.1 GI:472880
KEYWORDS anthranilate synthase I; anthranilate synthase II; trpE gene; trpG
gene.
SOURCE Buchnera aphidicola.
ORGANISM Buchnera aphidicola.
Bacteria; Proteobacteria; gamma subdivision; Buchnera.
1 (bases 1 to 4753)
REFERENCE
AUTHORS Lai, C.Y., Baumann, L. and Baumann, P.
TITLE Amplification of trpEG: adaptation of Buchnera aphidicola to an
endosymbiotic association with aphids
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (9), 3819-3823 (1994)
MEDLINE 94224833
REFERENCE
AUTHORS Lai, C.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1993) Chi-Yung Lai, Microbiology, University of
California-Davis, Davis, California, 95616-8665, USA
FEATURES
source
Location/Qualifiers
1..4753
/organism="Buchnera aphidicola"
/db_xref="taxon:9"
1..1728
/note="This 1.7kb EcoRV fragment is cloned separately and
not necessary contiguous with the rest of the sequence."
1..3580
/note="One unit of a four-unit plasmid"
1027..4753
/evidence="experimental"
/gene="trpE"
1027..2574
/gene="trpE"
/codon_start=1
/transl_table=11
/product="anthranilate synthase, component I"
/protein_id="CAA79931.1"


```

/db_xref="PID:q472881"
/db_xref="GI:472881"
/db_xref="SWISS-PROT:P42387"
/translation="MSKNPYEIEIIQKTAPEYHPDPTMAFNHLCASRPGTLLLETAENV
KRDESIIMIDSAMRISSEDSVKLTPISINGTIDILSTLTKTIKPIKLEIYKNNSTI
LVEPIKKNIDEDKLFSLVDFADFLMRIFENNEKSKNAFPGGLFSDIYSEES
LPLKGNKCSNCFYLAETLLVLDHOKTKCLIONSLEPSKNEKERRIKKREVEIER
LNEKLIKPKTKIKDINLTSNMNFEYCGITIKOKLIKOKIPIOVVSRRKYIECPN
PLSAQOKLKSNSPSPMPFMQODPTLTPASPESSSLKIDERTKTELPIAETPRGK
TGDNDLDDLSRIEEMKTNKELAEHLMLVDLARNDLARICKDSRYSDLVADR
YSHVHLVSRIEELREGDADLAHAYASCMNMGTLGAPVRAMOLIAEHEGERSYSG
GAIGYFTDGNLDTCTITRSAYVEKQATIQAGIVYNSIPENVENSLNAQAVIN
AIKNAHY"
2592..3182
/gene="trpE"
/codon_start=1
/translation="anthranilate synthase, component II"
/protein_id="CAA79932.1"
/db_xref="GI:472882"
/db_xref="PID:q472882"
/db_xref="GI:472882"
/db_xref="SWISS-PROT:P42388"
/translation="MANILLDNDSDFTYNNVEQDNRNNSNVLIRNTVDINTLINSI
KRNPILMLSPGSPSTPKNAGCMNLKIKKGEIPIVGLCHQAVAYAGIITAG
EIFHGKASLIHMDGLEMEFGLPQPLPARYSHLICKNKPKNILINSYENDMIMSRNN
LDVCGFOPHPSIILTTSALLLEKIIIMASIKYKXG"
4605..>4753
/gene="trpE"
/codon_start=1
/translation="anthranilate synthase, component I"
/protein_id="CAA79933.1"
/db_xref="GI:472883"
/db_xref="PID:q472883"
/db_xref="GI:472883"
/db_xref="SWISS-PROT:P42387"
/translation="MSKNPYEIEIIQKTAPEYHPDPTMAFNHLCASRPGTLLLETAENV
KRRL"
BASE COUNT      1949 a      615 c      599 g      1590 t
ORIGIN
Query Match      100.0%; Score 11; DB 1; Length 4753;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CCGTTCTAGT 11
      1111111111
      4725 CCGTTCTAGT 4715
RESULT 9
LOCUS      BS16S138      970 bp      mRNA      BCT      08-SEP-1997
DEFINITION      Bacterium sp. 16S ribosomal RNA, clone 400-138.
ACCESSION      AJ001424
VERSION      92385346
NID      AJ001424.1 GI:2385346
KEYWORDS      16S ribosomal RNA; 16S rRNA.
SOURCE      unidentified bacterium.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 970)
AUTHORS      Zaychikov E.F.
TITLE      Direct Submission
JOURNAL      Submitted (07-SEP-1997) Zaychikov E.F., Laboratory of Molecular
Enzymology, Limnological Institute, Ulan-Batorskaya str. 3,
Irkutsk, 664033, RUSSIA
REFERENCE      2 (bases 1 to 970)
AUTHORS      Denissova L.Y., Belkova N.L., Tulokhonov I.I., Parfenova V.V. and
Zaychikov E.F.
TITLE      Phylogenetic analysis of aquatic bacteria of the lake Balkal
JOURNAL      Unpublished
FEATURES      Location/Qualifiers

```

```

SOURCE      1..970
            /organism="unidentified bacterium"
            /db_xref="taxon:2338"
            /clone="400-138"
            <1..>970
            /gene="16S rRNA"
            /product="16S ribosomal RNA"
            /gene="16S rRNA"
            /gene="16S rRNA"
BASE COUNT      229 a      233 c      311 g      196 t      1 others
ORIGIN
Query Match      100.0%; Score 11; DB 1; Length 970;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CCGTTCTAGT 11
      1111111111
      Db      575 CCGTTCTAGT 585
RESULT 10
LOCUS      BSUB0003/c      209100 bp      DNA      BCT      26-NOV-1997
DEFINITION      Bacillus subtilis complete genome (section 3 of 21): from 402751 to
611850.
ACCESSION      299106 AL009126
VERSION      92632653
NID      299106.1 GI:2632653
SOURCE      Bacillus subtilis.
ORGANISM      Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE      1 (bases 1 to 209100)
AUTHORS      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertsch M.G., Bessières P., Bhotla A., Borchert S.,
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C.,
Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V.,
Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J.,
Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D.,
Emerson P.T., Entlan K.D., Errington J., Fabel C., Ferrari E.,
Fougeron D., Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A.,
Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J.,
Grandi G., Guisepi G., Guy B.J., Haga K., Hatach J., Harwood C.R.,
Henauf A., Hilbert H., Holsappel S., Hosono S., Bullo M.F.,
Itaya M., Jones L., Joris B., Karamata D., Kasahara Y.,
Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P.,
Koningstein G., Krog S., Kumano M., Kurita K., Lapidus A.,
Lardinois S., Lauber D., Lazarevic V., Lee S.M., Levine A., Liu H.,
Masuda S., Maueil D., Medigue C., Medina N., Meliado R.P.,
Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M.,
Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M.,
Portetle D., Porvolik S., Prescott A.M., Presecan E., Pujic P.,
Punelle B., Rapoport G., Ray M., Reynolds S., Rieger M.,
Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T.,
Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J.,
Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tanakoshi A., Tanaka T., Terresta P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
Viari A., Wambuit R., Wedler E., Wedler K., Welterger T.,
Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H. and
Danchin A.
TITLE      The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
JOURNAL      Nature 390 (6657), 249-256 (1997)
MEDLINE      98044033
REFERENCE      2 (bases 1 to 209100)
AUTHORS      Kunst F., Ogasawara N., Yoshikawa H. and Danchin A.
TITLE      Direct Submission

```


JOURNAL

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: mosze@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES

```

source      Location/Qualifiers
            1..209100
             /organism="Bacillus subtilis"
             /strain="168"
             /db_xref="taxon:1423"
             /gene="ycxa"
             /gene="ycxa"
             /gene="ycxa"
             /gene="ycxa"
             /function="unknown"
             /codon_start=1
             /transl_table=11
             /protein_id="CAB12161.1"
             /db_xref="PID:e1182320"
             /db_xref="PID:g2632654"
             /db_xref="GI:2632654"
             /db_xref="SWISS-PROT:008791"
             /translation="MRTSPRSKRVFLEFFVPAIGNSPNSFOFEMLPMAADAFADR
             SLISVSVIFMTTGTIOPEVGFIDRESVKIMAGVCISASFLVLPSPINAVFS
             AIVGVGGIGVSCAVGTTOYFISCMEDPHKGLALAILNANAGLVSPPIMAAP
             VYAGWOSTYITIGTMAAVLPVPLVFGMKHPHAQETKRTYDWGRGNVKKQSLI
             HILTVFTGCTGIGIDNAHLVPLINKDAVSVNMMAGFAFIIGGLAWMLDDL
             GSVYMLSLFELRLSLCLPLIGLHSDLMWFGFLFGLFGLSYGVIPPLTASIS
             ESYQVTLIGSLGILNFIHQVAGLSVAGLGFMTGCTGLIIVAVCIYFVGLSAVIE
             LVPLDKOKAKETHHSI"
             1280..1304
             /gene="ycxa"
             /complement(1292..1849)
             /gene="ycxb"
             /complement(1292..1849)
             /gene="ycxb"
             /function="unknown"
             /codon_start=1
             /transl_table=11
             /protein_id="CAB12162.1"
             /db_xref="PID:e1182321"
             /db_xref="PID:g2632655"
             /db_xref="GI:2632655"
             /db_xref="SWISS-PROT:008793"
             /translation="MIOYASESINLPGETTERDVEIEFFYOIAKISCFYFLFCAIFA
             AVNFINGPRIYVGSDALFNPNMILITMSTLFTLLLLLYVFSRAKKNEKMS
             KRTYTLNDEGIRICSKKYDLIFNDEITAVEYKNIFRVNTSSGOYIAPKHFSEEE
             EMNFEKILIKNETEKRLKFKKDH"
             complement(1903..2841)
             /gene="ycxc"
             /complement(1903..2841)
             /gene="ycxc"
             /function="unknown"
             /note="similar to hypothetical proteins from B. subtilis"
             /codon_start=1
             /transl_table=11
             /protein_id="CAB12163.1"
             /db_xref="PID:e1182322"
             /db_xref="PID:g2632656"
             /db_xref="GI:2632656"
             /db_xref="SWISS-PROT:008794"
             /translation="MYKHOQTPATYIAAILYSPITLSLEFKIAIQTAEPEDILAHNR
             TIAFAATVPLIFGVKLSILKVDILPLALPLALFESFOAGLVYSSSESGIIL
             OAAIPFTMVAAYLVKERPTWTOGFTVLVAVGWFVFKGVDVESAKGSLIL
             LSAISSAMNTAARMTQRFKLELTLYLSAIGVFNALVLRHGAAGTGTVEFLP
             RPPGVLAIVLGVSIVTSFLSNYTSRIEAFKGSAPFNHSTIVTNIAGFVILNES
             LAMVHLAGVCMIGVGSNINLEKTKRPGPAK"
             2965..4299
             /gene="ycxd"
             /gene="ycxd"
             2965..4299
             /gene="ycxd"
             /function="unknown"
             /note="similar to transcriptional regulator (GntR family)"

```

```

/aminotransferase (Mocr-like)"
/codon_start=1
/transl_table=11
/protein_id="CAB12164.1"
/db_xref="PID:e1182323"
/db_xref="PID:g2632657"
/db_xref="GI:2632657"
/db_xref="SWISS-PROT:008792"
/translation="MEKYSLSLTRIEEMQSTSYOEGDRLPSIROLARSYOVSKSTVI
RALQELERKHLIYVPSKSGYIVKSKGSKSGOPIDFATSAPDPVFPYLDPOHCI
KNAIDYKNDLEIYGTGPKLSILVRLKLLATOOVFDERHIFITSGVOQLSLICA
MPEPKREKIAIEOPGYHLMEULETIGIPVIGVRETEGDIARVETLFOESTKFF
YTMRFNHPICGCSERDKOELVRAEYDVIVYEDDILGDEENKRAKDPATAYLSS
HYTLKSTSKMFPGLRVGAALVPALTDGTTFYARKLNDIDCSMTSQALETYLSGM
YGRHEKIRDSYKERSLRHQAIRTRQLGSGRFTFSSGOAPCMHTLVLPDPLPASR
VIHRLKQGVLEAIDRHYLSDPENKLKINISVKTEDIERGKILMSHL"
complement(4472..4969)
/gene="gfp"
complement(4472..4969)
/gene="gfp"
complement(4472..4969)
/gene="gfp"
/codon_start=1
/transl_table=11
/product="surfactin"
/protein_id="CAB12165.1"
/db_xref="PID:e1182324"
/db_xref="PID:g2632658"
/db_xref="GI:2632658"
/translation="MKIYGIANDRPLSOENERPMTFISPKREKCRFRYKEDARSD
LIGDLYSVISROYOLDKIDLRSTOBYGKPCJPDLPDAFNTSHSGRWATGAFDSO
PIGIDIEKTRISLEIARFESKTEISDLKDKDEQTDITYHLSMKESITKQGROR
LIASA"
complement(5053..5071)
/gene="ycze"
complement(5053..5721)
/gene="ycze"
complement(5074..5721)
/gene="ycze"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12166.1"
/db_xref="PID:e1182325"
/db_xref="PID:g2632659"
/db_xref="GI:2632659"
/db_xref="SPTREMBL:034927"
/translation="MKOELVIRMTFFYAGLIIIAFGVSLTEGKALGISPDADHYSL
FOHGLYVGOSIITGALIVFTSLFTRAPKIGALNMLVIGVIFDENFILPAPST
YTGSIIVFSIGLVILIGVGVYVAGAGGRDSLMMLITEKTMNVQWANGMEITI
LFAAMGGGPIGFQGITLAILTGILIFLSPOSITLNYAVARRTAAVKASPVH"
complement(5815..6782)
/gene="ycxi"
complement(5815..5830)
/gene="ycxi"
complement(6039..6782)
/gene="ycxi"
/function="unknown"
/note="similar to glutamine ABC transporter (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB12167.1"
/db_xref="PID:e1182326"
/db_xref="PID:g2632660"
/db_xref="GI:2632660"
/db_xref="SWISS-PROT:P39456"
/translation="WLVKGLNKSFGENILIKRIDMKIERKVAIYIILGPGSGCTTLL
RCNLALIEIPRNGELAFDDFSIDSKVKVQADILKLRKSPNVQAVHLPFRALENV
MEGPVQVQKRKREVERKEAIOILDKVGILKMDLPPOLSGGQOQVGIARALAIQPE
LMUDEPTSLADPELVGEVLKVIDLANEGMTWVTVHEIKFAQVADDEVIFIDGVI
VEQGPPOIFSAPKEERTFORILNLNPL"
complement(6796..7500)
/gene="ycxj"

```


CDS
 complement(6796..7500)
 /gene="yckJ"
 /function="unknown"
 /note="similar to glutamine ABC transporter (permease)"
 /codon_start=1

Query Match 100.0%; Score 11; DB 1; Length 209100;
 Best Local Similarity 100.0%; Pred. No. 1,1e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR 1 CTTCTTCTACT 11
 |||||
 DB 171756 CTTCTTCTACT 171746

RESULT 11
 BS080011/c
 LOCUS
 DEFINITION
 .SSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Bacillus subtilis
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
 Bacillus subtilis
 1 (bases 1 to 207730)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
 Azevedo, V., Bertero, M.G., Bessieres, P., Bojotin, A., Borchert, S.,
 Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
 Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
 Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
 Daniel, R.A., Denzot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
 Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
 Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
 Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligorsky, E.J.,
 Grandi, G., Gulseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
 Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
 Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
 Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
 Koenigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
 Lardin, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
 Masuda, S., Meisel, C., Medigue, C., Medina, N., Melado, R.P.,
 Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
 Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Paro, V., Pohl, T.M.,
 Portetlelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
 Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
 Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
 Scanlan, E., Schleich, S., Schroeter, R., Scottone, F., Sekiguchi, J.,
 Sekowska, A., Seror, S.J., Serrito, P., Shin, B.S., Soldo, B.,
 Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
 Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpilzer, P., Tognoni, A.,
 Tosato, V., Uchiyama, S., Vandenbol, M., Vanler, F., Vassarotti, A.,
 Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzengraber, T.,
 Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
 Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
 Danchin, A.
 The complete genome sequence of the gram-positive bacterium
 Bacillus subtilis
 Nature 390 (6657), 249-256 (1997)
 98044033
 2 (bases 1 to 207730)
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 Direct Submission
 Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
 Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
 68 89 48 Location/Qualifiers
 1..207730

gene
 terminator
 CDS
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 complement(10..1436)
 /gene="yocA"
 complement(10..33)
 /gene="yocA"
 complement(45..1436)
 /gene="yocA"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13730.1"
 /db_xref="PID:e1185310"
 /db_xref="PID:g2634231"
 /db_xref="GI:2634231"
 /translation="METSQEPVLTQKSRFSVRSMSLFLVPLLSNVYQSVCQLVGM
 MANGKMLGDVAVAVSSFPFLFLSLFTIGSGSSILIGQAVGAKNEERLAVGT
 TLFTFLGLVAVLAVIGSIFLIDLRLMGPEPENVIHVSANARILFTAMPMPFLYAVT
 TFLRGDSKTPFTYLLIVSTVINIALPVLLLMGEPKLGIGSAVAVYISIAFL
 VLVVYLRKRHPLOPDFKTVRRYLKMKELVLLRLGVPASIMMIVLSIIVSVF
 NHYGSNATAVAGVNVQVAVYQVMPAVSIGAVSIFPAQSGANEPRLKQVITVGH
 NVIGCVLLITLVFVSHQILSLFTQDESILYHRLMTTLMSTYLLFGNQIISATMR
 ASGVLPYVVISIFALMGVEVPVAFVLSHTKLELLGVWGVPAFAVSVLLIYGYQ
 FVKKKQIITRLIO"
 1722..2345
 /gene="yocB"
 1722..2267
 /gene="yocB"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13731.1"
 /db_xref="PID:e1185311"
 /db_xref="PID:g2634232"
 /db_xref="GI:2634232"
 /db_xref="SPTREMBL:O34841"
 /translation="MKKCLFLTTLILSLSTNAPAKNTSGDLSQKALQALASARE
 HEMTSGNPNPKVAKVACSPGFEYONLQYVYMCSDLGKARAVNYLTFTPTAREK
 GPDVYFTYSKGLAIPIDGDMLLWKKSTAKLISKSGSTIITETVFTLDSPSAK
 RKTFTYKKNKMNQFDAYI"
 2274..2292
 /gene="yocB"
 2319..2345
 /gene="yocB"
 2359..2434
 /gene="trnS-Arg1"
 /product="transfer RNA-Arg"
 2359..2434
 /gene="trnS-Arg1"
 complement(2745..3032)
 /gene="yocC"
 complement(2745..3032)
 /gene="yocC"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13732.1"
 /db_xref="PID:e1185312"
 /db_xref="PID:g2634233"
 /db_xref="GI:2634233"
 /db_xref="SPTREMBL:O35009"
 /translation="MHIYQIRSLERIOEVKQYLLKNKRQDYFLFIQNSALRISDI
 LPLOYKQVONKDHVATSONEKKKDSFRFAEAGNIRYVRRHGRKRSF"
 complement(3327..3347)
 /gene="yocD"
 complement(3327..3578)
 /gene="yocD"
 complement(3348..3578)
 /gene="yocD"
 /function="unknown"

Query Match

```

/codon_start-1
/protein_id="CAB13733.1"
/db_xref="PID:e1185313"
/db_xref="PID:g2634234"
/db_xref="GI:2634234"
/db_xref="SPTREMBL:O34555"
/translation="MTITETETATNTLSDYIKSLVLEKIRAVHDEQFLYKKEQF
KTHLEQLENTKALVQETLNEPIPEDIDVKED"
/gene="5526
3763..5526
/gene="ygf"
/ec_number="2.3.2.2"
/function="glutathione metabolism"
/note="alternate gene name: pac"
/codon_start-1
/transl_table-11
/product="gamma-glutamyltranspeptidase"
/protein_id="CAB13734.1"
/db_xref="PID:e1185314"
/db_xref="PID:g2634235"
/db_xref="GI:2634235"
/db_xref="SPTREMBL:O34555"
/translation="MKRTWVCLTALLSVLVAGSVFHAANKPKRSDYKQVDVG
KQDVATAPHLAEIGADYLVKGNDAIDAVALNTQFALNTEPMKSGIGGGFMYVDG
KTKDTTIDISREBPAGAPDMFLDENGKAIPESEYTKGTAVGPGTLEGEALDK
WGRSMKOLITPITKLAEKGPIDSVLAISDYERLSTPAKDVPIGELKSGD
TLQKDLATFKLIRSGKDAFKKFKAKTSLTYODPFGSWEKYLENDITIDEI
WQDYQYQITTPPSSGGIFLLQMKILDFNLSDYVRSWEKYLENLEMTLSYAD
KASYADPEVNVPLGLHPDIKERQQLINDOVNKKRKADPKRYOGESANYQAP
EOPKDEQGTHTFVADRWGNVSYTTTEQLFGTGIMVDPYGVILNLETFDPAIP
GGNEVQPNKRPKPSMTPTLKDKPVLTVGSPGATITISVLOTILHIEKMEK
AAVEEPRTYNSMSYRYEDGVPKDVLKNGMNGHFGSPVDIGNVQISIDHENGIT
FKGVADSSNGAIGINTK"
/gene="5535..6484"
/complement(5535..5559)
/gene="yofa"
/gene="yofa"
/complement(5627..6484)
/gene="yofa"
/function="unknown"
/note="similar to transcriptional regulator (lysr family)"
/codon_start-1
/transl_table-11
/protein_id="CAB13735.1"
/db_xref="PID:e1185315"
/db_xref="PID:g2634236"
/db_xref="GI:2634236"
/db_xref="SPTREMBL:O34685"
/translation="MESGDLKIFQAVARSGSTIKAKOMLNVYQNTARVHNEEDLN
IRLFHRTNGMKLTLAGENLLOYADQVLSILOAKERTSKPRKPLIGSLETNAY
THLPEHASFIRRFPEVDVSYNADTNNHILQOYLDKRYGATYGVHEAARQQLVS
HBEVLITTSREGAEMLQOPMFLFEGASHRDVNRVRLLEAGIHOKKITEGTLEAI
INGVSAGMTALLPKSADVSGSEHRTVMWHLPPSYQDEIVFIYKDKDFITSAFQTE
LDEINEMKR"
/gene="7601
6612..7601
/gene="yofa"
/gene="yofa"
/function="unknown"
/note="similar to alcohol dehydrogenase"
/codon_start-1
/transl_table-11
/protein_id="CAB13736.1"
/db_xref="PID:e1185316"
/db_xref="PID:g2634237"
/db_xref="GI:2634237"
/db_xref="SPTREMBL:O35017"
/translation="MKAVTHNGKAGLITGVQDVPSTKRGVEYVKLSAGLNHRDL
FLMKKSEDDPHMILGSDAGITIEEGEVKANTVOTEVVIFTLNMDLTENVPPVE

```

100.0% Score 11: DB 1: Length 207730:

Best Local Similarity 100.0%: Pred. No. 1.2e-03:
Matches 11: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
y 1 CTGTCTCTAG 11
db 84140 CTGTCTCTAG 84130

RESULT 12
BSUB0016/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.

REFERENCE 1 (bases 1 to 215640)
AUTHORS
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertsche, M.G., Bessieres, P., Bolo, A., Borchert, S.,
Boriss, R., Boursier, L., Brans, A., Brun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerth, I.F., Cummings, N.J.,
Daniel, R.A., Denicot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entlian, K.D., Errington, J., Fabre, C., Ferrari, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fume, S., Gallizzi, A.,
Galleron, N., Gilm, S.Y., Glaeser, P., Goffeau, A., Golligly, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Hags, K., Haele, J., Harwood, C.R.,
Henatt, A., Hilbert, H., Holtsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerner-Blanchard, M., Klein, C., Kobayashi, Y., Koeltter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A.,
Masuda, S., Mauch, C., Medina, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Paro, V., Pohl, T.M.,
Portetle, D., Portetle, S., Prescott, A.M., Prescott, E., Pujic, P.,
Purcell, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadleir, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scrofer, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Seror, P., Shin, B.S., Solio, B.,
Sorokin, A., Taconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassart, A.,
Viarl, A., Wambuit, R., Wedler, E., Wedler, H., Weitzenecker, T.,
Winers, P., Wipat, A., Yamamoto, H., Yamane, K., Yasunoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

TITLE The complete genome sequence of the gram-positive bacterium
Bacillus subtilis

JOURNAL Nature 390 (6657), 249-256 (1997)

MEDLINE 98044033

REFERENCE 2 (bases 1 to 215640)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submissio

AUTHORS Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES Location/Qualifiers
1..215640
/organism="Bacillus subtilis"
/strain="168"

terminator
/db_xref="taxon:1423"
complement(8..24)
/gene="ytmn"
complement(8..1001)
/gene="ytmn"

gene


```
CDS      complement(99. .1001)
          /gene="ytmn"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14906.1"
          /db_xref="PID:e1185801"
          /db_xref="PID:g2635412"
          /db_xref="GI:2635412"
          /gene="SPTREMBL:034430"
          /translation="MKRIEAFIFGLISQIDSLGMAVGVSTSLIAFGITPAVAS
          ASYLAEVVTTAGSVGSHKEFNGVNDKOTVOIVIPGSIAGFAGATLSOLPGVAPY
          ISFILLGLGVILIRLFQYKPALEKHHVPLNKKQSIPLGVITGFPADATSGGNGVY
          TPIILSRKGLSPKRVGVDTSEFALVASTAGFLISLGMEDVNMVLSLWAGGIIA
          APIAAVYOKFHPDLMGVLVGCFILVNARTLINEMVANTAVHPLIYTAIGAIWLSAV
          LFVLSKIGNRIYKTAVDVHLKEK"
          complement(1020. .2270)
          /gene="h1p0"
          complement(1020. .2270)
          /gene="h1p0"
          /EC_number="3.5.1.32"
          /codon_start=1
          /transl_table=11
          /product="hippurate hydrolase"
          /protein_id="CAB14907.1"
          /db_xref="PID:e1185802"
          /db_xref="PID:g2635413"
          /db_xref="GI:2635413"
          /db_xref="SPTREMBL:034980"
          /translation="MSLDYRNIEGSPYOTGNDILTLKEESPVNLSTLEKOLIGI
          RRLHOPETSKKEEPTAFIRKCKEKQGIQITPALKTVFADIDGESGPIALARA
          DIDALPIEKTGLPIASKHKGIVHAGCHDEHIALGAAFLKENDSLEKGRLLFQ
          PAEAGGATKIVIEDQDIDAVIGLHNKPDIAVGTGKTPLAADVREVELEG
          KGAAALPHNGEDPIIGASQILVALQIVSRNPNQOSALITVGKINGSTWVVIDT
          VLEGVTRTFDSEVRNQKRFPAVTEIOISAPSLKANYHSGPPPLCNDDEITGLV
          RDAARAKILOVIDPASTAGEDFAYYLEHIPSFAFFGTDGDHMHHPATIDETAILI
          KASYFLYESAKRLIDSNESKISD"
          complement(2288. .2980)
          /gene="r1br"
          complement(2288. .2980)
          /gene="r1br"
          /function="regulation of riboflavin biosynthesis genes"
          /note="alternate gene name: ytmk"
          /codon_start=1
          /transl_table=11
          /product="transcriptional regulator"
          /protein_id="CAB14908.1"
          /db_xref="PID:e1185803"
          /db_xref="PID:g2635414"
          /db_xref="GI:2635414"
          /db_xref="SPTREMBL:034965"
          /translation="MTIIAGTVVKGOLGKLGPTANVDARIHGLRNGVYGLAVTN
          HOHLAMINIGVPTVGSNLEKTELEIFEPHDIYGEKIEGSLFKRIEERPDLE
          SLRKOIKDISCVAKRELIGMAPKKESLSHOLELCTCFYKCKNIXYVNGV
          YNVIDNPEFYGTQVAYRRIYILSFLSKEDNPVSSKTIIFGAGLADLNKRITS
          SIYEESEENILG"
          complement(3027. .4355)
          /gene="ytmj"
          complement(3027. .4355)
          /gene="ytmj"
          /function="unknown"
          /note="similar to nitrolicetate monoxygenase"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14909.1"
          /db_xref="PID:e1185804"
          /db_xref="PID:g2635415"
          /db_xref="GI:2635415"
          /db_xref="SPTREMBL:034974"
          /translation="MTRADFIQFGAMHGVGTDDGMRHPDVPSASTNIEFYKKAQ
          TAEKGFSEFIFADGLFISEKSIPIHFNREPLIIISALASVTKNIGVGTSTSTE
          PTISROIWSDIHSGRAGMNLVTSPOEAAHNKSNLPEHTEYETIAOEHLVDYR
          GLNWSNHDPAFINNKITGCFDQAKLHRLNKKITFOVEGPLNIGSKOGEPPVFOAG
```

```
SEETGRQPAKNAADALFTHSNSLEETKATYADYKSRADGCDPSSVRIFPCISPIVA
DTEERAKRYREFAEILIPLENAVYLAERFDYDLSVPLDPEFPDIDGVGNAAOST
TDRIKREAKARNLTLREVAOEMAFRTLEIGPEVVASLIEFWFNAADGFIWGSIDI
PGTIDAEVKEVPILOERGLYRDYRGGLFRNLGIGIQHOSVLAHSSH"
          complement(4352. .4633)
          /gene="ytmj"
          complement(4352. .4633)
          /gene="ytmj"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14910.1"
          /db_xref="PID:e1185805"
          /db_xref="PID:g2635416"
          /db_xref="GI:2635416"
          /db_xref="SPTREMBL:034639"
          /translation="MSDVNIYVMSKKGSGVCEYKKNYNENKGFPONDISEKELR
          DILQYKGYGRHVPVYEIGRNOYQGITTEIGIEHLDLANHQAIREAKR"
          complement(4648. .5652)
          /gene="ytm0"
          complement(4648. .5652)
          /gene="ytm0"
          /function="unknown"
          /note="similar to hypothetical proteins"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14911.1"
          /db_xref="PID:e1185806"
          /db_xref="PID:g2635417"
          /db_xref="GI:2635417"
          /db_xref="SPTREMBL:034846"
          /translation="MIRSLIDQSLIGEGETAADTLQHTVKLAQNAEEGVNRFVAE
          HNNDEIAGSABEVLGLYLASTRKIRLASGVMQHSYVAAQFHLISALAGRI
          DGNVKAQPGFQSLDLOAEYKRPVRODELEETHEVDPDTHRYALRSPQ
          VDRKGIFLIGSTESASIAKLGISFVAYVINGEELPKARAPAHUAPGSEAE
          FLAAVAVPAHREKREKHYVRESIKYVLKGRKRVNNGSRQAAHYLENTEPIDIT
          YOKTGIAGTKEEVAELRLSGTYKINDVIFTPIKNVEKOLSTYLLSDVLAAR
          "
          complement(5649. .6428)
          /gene="h1sp"
          complement(5649. .6428)
          /gene="h1sp"
          /function="histidine transport protein (Atp-binding
          protein)"
          /codon_start=1
          /product="histidine transport protein (Atp-binding
          protein)"
          /protein_id="CAB14912.1"
          /db_xref="PID:e1185807"
          /db_xref="PID:g2635418"
          /db_xref="GI:2635418"
          /db_xref="SPTREMBL:034900"
          /translation="MIEIKNIKRGSHHVLKGINLTVRKGEVYTIIGPSGSKTFL
          RLNLLEPDEEIIISHDKYINCRSPSKEVNMLKQOTAMPQOHLPAHAKVINYNA
          BEITLARKKRODAVAVENEDRKGLDKNLAIYPSQSGGQKGVGLARALAIHPDV
          LLEDFPDLADBELVGEVLELVKGTATVIVTHMEFARRRSDOVFEMDEGVIV
          EOGTPEVEYRHKRKDRTOFLRVRVEYLFEPEKRIKEPVI"
          complement(6425. .7132)
          /gene="ytmw"
          complement(6425. .7132)
          /gene="ytmw"
          /function="unknown"
          /note="similar to amino acid ABC transporter (permease)"
          /codon_start=1
          /transl_table=11
```

```
Query Match      100.0%; Score 11; DB 1; Length 215640;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CTTGTTCTAGT 11
Db 189460 CTTGTTCTAGT 189450
```


RESULT 13
 BTPICRGEN/C
 LOCUS BTPICRGEN 1707 bp DNA BCT 25-JUL-1996
 DEFINITION B. thuringiensis pICR gene and orf2.
 ACCESSION X93361
 NID X93361.1 GI:1212804
 VERSION X93361.1 GI:1212804
 KEYWORDS ORF2; pICR gene; pICR protein.
 SOURCE Bacillus thuringiensis.
 ORGANISM Bacillus thuringiensis
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.

REFERENCE
 1 (bases 1 to 1707)
 Lereclus, D., Agaisse, H., Gontnet, M., Salameau, S., and Sanchez, V.
 Identification of a Bacillus thuringiensis gene that positively
 regulates transcription of the phosphatidylinositol-specific
 phospholipase C gene at the onset of the stationary phase
 J. Bacteriol. 178 (10), 2749-2756 (1996)
 96212013
 2 (bases 1 to 1707)
 Lereclus, D.
 Direct Submission
 Submitted (22-NOV-1995) D. Lereclus, Unite de Biochimie
 Microbienne, Dept des Biotechnologies, Institut Pasteur, 28 Rue du
 Dr. Roux, 75724 Paris Cedex 15, FRANCE
 3 (bases 1 to 1707)
 Direct Submission
 Submitted (01-MAR-1996) D. Lereclus, Unite de Biochimie
 Microbienne, Dept des Biotechnologies, Institut Pasteur, 28 Rue du
 Dr. Roux, 75724 Paris Cedex 15, FRANCE

REFERENCE
 AUTHORS Lereclus, D.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-1996) D. Lereclus, Unite de Biochimie
 Microbienne, Dept des Biotechnologies, Institut Pasteur, 28 Rue du
 Dr. Roux, 75724 Paris Cedex 15, FRANCE

FEATURES
 source
 1..1707
 /organism="Bacillus thuringiensis"
 /strain="407"
 /db_xref="taxon:1428"
 418..467
 515..1372
 /gene="pICR"
 515..1372
 /gene="pICR"
 /function="positively regulates transcription of
 phosphatidylinositol-specific phospholipase C gene"
 /codon_start=1
 /transl_table=11
 /product="pICR protein"
 /protein_id="CA63744.1"
 /db_xref="PID:e214719"
 /db_xref="PID:g1212805"
 /db_xref="GI:1212805"
 /db_xref="SPRMBL:045782"
 /translation="MOAKLGSPIKIRIVRGITOKOLSENICHOSEVSRISGAVNP
 SMDLIGIAKLIPIITHFREVLIDIEKKQFQOVIMLCOKRKIKITNVNML
 KKEVYPERQQLQWQYVAAVYALKVDYECLEIKLNDLTGIDYONLYINA
 IANIVANGELKKQIDLEFQILKQLEALDNEEDVAVYNHAKALYLDSEESLYQ
 VNKAIEISCRINSMALIGOLYVQREGCLRLVEEAIEDAVYKKAFFDILEMAYK
 EALVNKISR"
 1461..1607
 /gene="orf2"
 /gene="orf2"
 1461..1607
 /gene="orf2"
 /codon_start=1
 /transl_table=11
 /protein_id="CA63745.1"
 /db_xref="PID:e224565"
 /db_xref="PID:g1212806"
 /db_xref="GI:1212806"
 /db_xref="SPRMBL:045783"
 /translation="MKKLLIGSLITLAMAAGISLGDTAFKSIISHNDQEVVAADL
 PPEF"

BASE COUNT 614 a 203 c 321 g 569 t

ORIGIN

Query Match 100.0%; Score 11; DB 1; Length 1707;
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
 Db 1676 CTGTTCTACT 1666

RESULT 14
 BUHRDDG/C
 LOCUS BUHRDDG 4877 bp DNA BCT 22-JUN-1994
 DEFINITION Buchnera aphidicola ribosomal inner membrane protein
 (rnpA-rmpH-dnaA-dnaN-gyrB) gene, complete cds.
 ACCESSION M80817
 NID M80817.1 GI:144144
 VERSION M80817.1 GI:144144
 KEYWORDS dnaA gene; dnaN gene; gyrB gene; inner membrane protein; ribosomal
 protein; rnpA gene; rmpH gene.
 SOURCE Buchnera aphidicola DNA.
 ORGANISM Buchnera aphidicola
 Bacteria; Proteobacteria; gamma subdivision; Buchnera.

REFERENCE
 1 (bases 1 to 4877)
 Lai, C.-Y. and Baumann, P.
 Genetic analysis of an aphid endosymbiont DNA fragment homologous to
 the rnpA-rmpH-dnaA-dnaN-gyrB region of Eubacteria
 Gene 113, 175-181 (1992)
 92241666

REFERENCE
 AUTHORS Lai, C.-Y. and Baumann, P.
 TITLE Genetic analysis of an aphid endosymbiont DNA fragment homologous to
 the rnpA-rmpH-dnaA-dnaN-gyrB region of Eubacteria
 JOURNAL Gene 113, 175-181 (1992)
 MEDLINE 92241666

FEATURES
 source
 location/Qualifiers
 1..4877
 /organism="Buchnera aphidicola"
 /db_xref="taxon:9"
 complement(1..821)
 /note="60 kDa"
 /codon_start=1
 /transl_table=11
 /product="inner membrane protein"
 /protein_id="AA73145.1"
 /db_xref="PID:g551760"
 /db_xref="GI:351760"
 /translation="MELQNFPIFAFLVPSFLWQAMOSQSLKRNKNEENSPFHLN
 HKREKNOIIRKNDVLRVINYGGDIEASLDYKQLNSPENKLEFSQDVEYQA
 OSLGKDGKDSINKRIPYSAKKNFELGHEKELRPIPIAKNGITIKNFIK
 RGVAVVEYDINNNSKRLKLENTGQAKQYKPKPNNDIYSGNALTFFGAASS
 DAKYKIKFDNIANHEHLHITTHGWMALQOYFVAVWPDNISSNITITSTL"
 complement(824..1081)
 /note="10 kDa"
 /codon_start=1
 /transl_table=11
 /product="inner membrane protein"
 /protein_id="AA73146.1"
 /db_xref="PID:g144146"
 /db_xref="GI:144146"
 /translation="MVKSTIVFCLFFSVSYONYISFMPNSNCRFYPTGTYMILS
 LRKREVGITITLIRLRCCHPLHGSGEDIVPKIKKSEY"
 complement(1060..1404)
 /gene="rnpA"
 /gene="rnpA"
 complement(1060..1404)
 /gene="rnpA"
 /codon_start=1
 /transl_table=11
 /protein_id="AA73147.1"
 /db_xref="PID:g144147"
 /db_xref="GI:144147"
 /translation="MLNVEFKKSKSLKSTNQVYFSNPCKNKFHINILGRSNLGH
 PLGSLSRKNKIKAVRNKIKRLIRETFPLQLHRLIMDQVYIAKKNIYVLLNNKIY
 NILEYINSNYOR"
 complement(1418..1561)
 /gene="rmpH"
 /gene="rmpH"
 complement(1418..1561)
 /gene="rmpH"

gene
 CDS


```

/codon_start=1
/transl_table=1
/product="ribosomal protein"
/protein_id="AA073148.1"
/db_xref="PID:9144148"
/transl_table="MKRTFOPSLIKRNSHGFRPMATKNGRYILSRRAKRLRLTV
SSK"
2049..3344
/gene="dnaa"
2049..3344
/gene="dnaa"
/codon_start=1
/transl_table=1
/protein_id="AA073149.1"
/db_xref="PID:9144149"
/transl_table="MWIRSLKALNNILLYANQFILDWVKDYLIHFKKILQDFC
GNSPFIKFKYOTSKERKFKKILQKONAKIPIDKIPFKSSHRNINKHSE
ENLEKGNOLARAASOVAKNPNSSNPLEYGTGTHLHAIGLIVAKYAV
KIYHSERFYQDMKALONNAIEKFLYYSVALLIDIOFPAHKEBSQEFFHTF
NALLEGNOQIILISDRPKKEINGVEDLKSIFGGLTVAIDPELETRVALILKAE
NIVLSEVAFIAKHLRSNVELEGALNRYIVANFTHSITVEFVREALDILAO
EKLVTIANIOKTVAEYKIKYADLSRRSRSVARPMAMAKELTNHSLPEIGDA
FSGDHTVLAACRKIEQLREKNDIKEDFNLIRLTSV"
3349..4449
/gene="dnaa"
3349..4449
/gene="dnaa"
/codon_start=1
/transl_table=1
/protein_id="AA073150.1"
/db_xref="PID:9144150"
/transl_table="MKFTIONDILTRNMLKTRVYLVKNISFPILENILIOVEDTSL
TTNLEILSLIKIEITIKYIPKTTISGRKILICRTISEKSIKMKQKKMYISE
NSNYILSLADTPHONFPIKSPDISNLIKEMTEKESKQDVYVYINGMLT
EKKDFRSVATDGYRLAISTOLKDNFSSITPKAYMELIKLNTVOPOLNITI
GNSIRITKTLITTOILIBEGYPIYKSVLFKEKKNITINSILKSLVAIHLAE
KFCGIEIKENGKRVLSDNOEETADLEIDYFGEKIEISINWYVLLDINNIXSE
NIALFLNKSRSIOIEANNSNAYVMLKR"
4604..4877
/gene="gyrB"
4604..4877
/gene="gyrB"
/partial
/gene="gyrB"
/codon_start=1
/transl_table=1
/protein_id="AA073151.1"
/db_xref="PID:9551761"
/db_xref="GI:551761"
/transl_table="MIDYDSKIKILRGDAVAKRPGMYIGDTDGSGLAHNVETIV
DNSIDELALGFCFEIKRVYIHSDNSVSKDGDGRGIPDTHPEKISAA"

BASE COUNT      1955 a      617 c      605 g      1700 t
ORIGIN

Query Match      100.0%; Score 11; DB 1; Length 4877;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CTGTTCTAGT 11
      |||||
DB      1434 CTGTTCTAGT 1424

RESULT 15
BUHTRPEA/c      3641 bp      DNA      BCT      14-AUG-1996
LOCUS      Buchnera aphidicola anthranilate synthase large subunit (trpG) gene
DEFINITION      and anthranilate synthase small subunit (trpG) gene, complete cds.
ACCESSION      L43551
ID      g1082002

```

```

VERSION      L43551.1 GI:1082002
KEYWORDS      anthranilate synthase; large subunit; small subunit.
SOURCE      Buchnera aphidicola DNA.
ORGANISM      Buchnera aphidicola
REFERENCE      1 (bases 1 to 3641)
AUTHORS      Roubaksh.D., Lai,C.Y., von Dohlen,C.D., Clark,M.A., Baumann,L.,
Baumann,P., Moran,N.A. and Voegelin,D.J.
TITLE      The tryptophan biosynthetic pathway of aphid endosymbionts
(J. Mol. Evol. 42 (4), 414-421 (1996))
JOURNAL      J. Mol. Evol. 42 (4), 414-421 (1996)
MEDLINE      96215866
FEATURES
     source
         location/Qualifiers
             1..3641
             /organism="Buchnera aphidicola"
             /specific_host="Rhopalosiphum padi"
             /db_xref="taxon:9"
             1..1545
             /gene="trpG"
             1..1545
             /gene="trpG"
             /codon_start=1
             /transl_table=1
             /product="anthranilate synthase large subunit"
             /protein_id="AA05986.1"
             /db_xref="PID:g1082003"
             /db_xref="GI:1082003"
             /transl_table="MKKSAYPIEIIIOKAPHPDPYMFNHNLCESRSETLLETADNK
KRKRKIMTIDSMRISSEHNAVKLPISINGMEITLVYKTIISKIEIYRRKKNIL
IPFNIRKQDEDEKLESYVDFARLITFERREKSKAMFFGGLSTDLSEITL
PKLGNQKSNCFEYLAETLVLDHQQKTLQSSFTNSNEKRIEERSEIKOL
NOYKLNIPKIKIDILNNSNNEFGSIIKIKLOKGEIIOVPSRKYPCPNP
LSAYOKLRKSNPSPIYFQDQDFTLFGASPESSLYDKTRIEIYPIAGTRPGRT
KQNDLDDSRLEEMRTNHELAELMLVLDARNDLARIKCPGSRYSDDLVRVRY
SHMHVSVSIVGELREGDLPALAHVYSCNMGTITGPKYCANQLIYERKRGSTGG
AICFTDNLNDLCITIRSAIYERKICYNQAGIYVNSIPEDEVNESLNKQAVINA
IKNAH"
             1561..2149
             /gene="trpG"
             1561..2149
             /gene="trpG"
             /codon_start=2
             /transl_table=11
             /product="anthranilate synthase small subunit"
             /protein_id="AA05987.1"
             /db_xref="PID:g1082004"
             /db_xref="GI:1082004"
             /transl_table="MADITLDNDSFTYNLVEQLKNNNNVLIYRNNIDIKIINSI
NKIKNPILSPGSPKKNAGCMENINVEKRIPIVIGICLHQAIVEAYGCIIGYAG
EIFHGRALINHDGLEMEFGLPOLPVARYHSLCNKIRKPNFINSYFMDMINSVANN
LDYVCGFOHPSPSITTSALLERKIDWASSKY"

BASE COUNT      1490 a      453 c      503 g      1195 t
ORIGIN

Query Match      100.0%; Score 11; DB 1; Length 3641;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CTGTTCTAGT 11
      |||||
DB      121 CTGTTCTAGT 111

Search completed: September 17, 1999, 22:01:12
Job time: 16:25 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:31 ; Search time 425.19 Seconds

(without alignments)
6.473 Million cell updates/sec

ALIGNMENTS

Title: us-08-956-518a-88

Perfect score: 11

Sequence: 1 CTGTTCTACT 11

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	11	100.0	N91020	Nicotiana tabacum
2	11	100.0	N93197	Fragment of clone
3	11	100.0	3763	Sequence encoding
4	11	100.0	3316	Lambda gt10ch101 e
5	11	100.0	3688	Lambda gt10ch201 e
6	11	100.0	2863	Lambda gt10ch2 enc
7	11	100.0	3824	Sequence encoding
8	11	100.0	1920	MSP gene in pUCRS
9	11	100.0	1592	Usp45 gene, lactoc
10	11	100.0	391	Human brain Express
11	11	100.0	2232	Human death associ
12	11	100.0	531	Human prostate-sp
13	11	100.0	5836	Gene encoding BRCA
14	11	100.0	5712	BRCA1, human breas
15	11	100.0	24026	BRCA1 gene nonsens
16	11	100.0	5914	BRCA1 gene framesh
17	11	100.0	5914	BRCA1 gene missens
18	11	100.0	5915	Human gene signatu
19	11	100.0	5914	Mutated BRCA1 codi
20	11	100.0	5903	Mutated BRCA1 codi
21	11	100.0	153	Mutated BRCA1 codi
22	11	100.0	5914	Mutated BRCA1 codi
23	11	100.0	5912	Mutated BRCA1 codi
24	11	100.0	5910	Mutated BRCA1 codi
25	11	100.0	5910	Mutated BRCA1 codi
26	11	100.0	5914	Mutated BRCA1 codi
27	11	100.0	5913	Mutated BRCA1 codi
28	11	100.0	5913	Mutated BRCA1 codi
29	11	100.0	5912	Mutated BRCA1 codi
30	11	100.0	5914	Mutated BRCA1 codi
31	11	100.0	24025	Mutated BRCA1 geno
32	11	100.0	5924	Mutated BRCA1 codi
33	11	100.0	5914	Mutated BRCA1 codi
34	11	100.0	5915	Mutated BRCA1 codi
35	11	100.0	5874	Mutated BRCA1 codi
36	11	100.0	5915	Mutated BRCA1 codi
37	11	100.0	5915	Mutated BRCA1 codi
38	11	100.0	5914	Mutated BRCA1 codi
39	11	100.0	5913	Mutated BRCA1 codi
40	11	100.0	5912	Mutated BRCA1 codi
41	11	100.0	5913	Mutated BRCA1 codi
42	11	100.0	5914	Mutated BRCA1 codi
43	11	100.0	5913	Mutated BRCA1 codi

C 44	11	100.0	5914	1	T17468	Mutated BRCA1 codi
C 45	11	100.0	597	1	X30620	H. pylori cellular

RESULT 1	
ID N91020/c	N91020 standard; DNA; 735 BP.
AC N91020:	
DT 21-JUN-1990	(first entry)
DE Nicotiana tabacum cDNA for pathogenesis-related (P-R) protein	
KW Pathogenesis-related protein; Nicotiana tabacum cv. Samsun NN;	
KM pNTSNNCPRI/81; pNTSNNCPRI/1.	
OS Nicotiana tabacum cv. Samsun NN.	
PH Key	Location/Qualifiers
FT misc-feature	1..722
FT	/*tag- a
FT	/note="cDNA from pNTSNNCPRI/1"
FT	14..735
FT	/*tag- b
FT	/note="cDNA from pNTSNNCPRI/81"
FT	1..497
FT	/*tag- c
FT	498..735
FT	/*tag- d
PN	EP-307841-A.
PD 22-MAR-1989.	
PE 12-SEP-1988; 114877.	
PR 15-SEP-1987; US-096579.	
PA (GEHO-) Gen Hospital Corp.	
PI Pfltzner UM, Pfltzner AP, Goodman HM;	
DR WPI: 89-087235/12.	
DR P-PSDB; P91058.	
PT Pathogenesis-related proteins and regulatory sequences	
PT used for expression in plants to enhance hypersensitive response	
PT to invading pathogen	
PS Figure 3: 33pp; English.	
CC It is cDNA from tobacco mosaic virus (TMV) infected tobacco plants.	
CC It contains the sequence from two independent cDNA clones	
CC pNTSNNCPRI/81 and pNTSNNCPRI/1. Differences between the two clones are	
CC shown in the Features table. Expression of a recombinant DNA sequence	
CC which includes a PR-1 promoter in a plant enhances the hypersensitive	
CC response of the plant to an invading pathogen.	
CC Sequence 735 BP; 227 A; 145 C; 145 G; 218 T;	

QY	1	CTGTTCTACT 11	
DB	377	CTGTTCTACT 367	

RESULT 2	
ID N93197	N93197 standard; DNA; 5406 BP.
AC N93197:	
DT 28-JUN-1990	(first entry)
DE Fragment of clone lambda HK65a contg. kallikrein gene.	
KW Kallikrein gene; vasodilator; male infertility; lambda HK65a; ss.	
OS Homo sapiens.	
PH Key	Location/Qualifiers
FT exon	803..848
FT	/*tag- a
FT	2538..2697
FT	/*tag- b
FT	3967..4256
FT	/*tag- c
FT	4375..4334
FT	exon


```

FT      exon                      /*tag= d
FT      4838..5075              /*tag= e
PD      EP-297913-A.
PN      04-JAN-1989.
PR      30-JUN-1988; 306039.
PR      30-JUN-1987; US-068594.
PA      (AMGE-) Amgen Inc.
PI      Lin FK, Lu HS;
PI      WPI; 89-009139/02.
DR      P-PSDB; P93722.
PT      New recombinant kallikrein polypeptide(s) and encoding DNA -
PT      useful as vasodilators and for treating male infertility.
PS      Table VI, page 13; 43pp; English.
CC      Lambda HK64a is a subclone produced from clones isolated from a CHA phage
CC      bourse human fetal liver genomic library using monkey kallikrein MM680a
CC      clone DNA as a probe.
CC      See also N93193-8.
CC      Sequence 5406 BP; 974 A; 1864 C; 1309 G; 1259 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 5406;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      3532 CTGTTCTAGT 3542

RESULT 3
Q06328/c
AC      Q06328;
ID      Q06328 standard; DNA; 3763 BP.
DE      06-FEB-1991 (first entry)
DE      Sequence encoding merozoite apical-end-localised protein (MAEP)
DE      Insert 5.3.
KM      Malaria; vaccine; Duffy blood group; ds.
OS      Plasmodium vivax.
PN      WO9011772-A.
PD      18-OCT-1990.
PF      03-APR-1990; U01849.
PR      05-APR-1989; US-334041.
PR      06-APR-1989; US-334270.
PA      (UYNV-) NEW YORK UNIT.
PI      Barnwell JM, Galinski MR, Wertheimer SP;
PI      WPI; 90-334616/44.
PI      P-PSDB; R07503.
PT      Malarial apical end merozoite proteins and peptide(s) - used for
PT      developing cpos. for treating, preventing and diagnosing malarial
PT      infection.
PS      Claim 2; Fig 1a; 66pp; English.
CC      A MAEP compound having a binding affinity for a Duffy blood group
CC      antigen of primate red blood cells, is antigenic for the complete
CC      protein, and may be used in diagnosis, treatment and vaccination
CC      against invasion by P.vivax and P.knowlesi.
CC      Sequence 3763 BP; 1796 A; 464 C; 611 G; 892 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 3763;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      2305 CTGTTCTAGT 2295

RESULT 4
Q05633
AC      Q05633 standard; DNA; 3316 BP.
ID      Q05633;
DE      17-DEC-1990 (first entry)

```

```

DE      Lambda gtl0ch101 encodes C-terminal alpha amidating enzyme.
KM      C-terminal alpha amidating enzyme; human thyroid gland;
KW      Lambda gtl0ch101; ss.
OS      Homo sapiens.
FH      Key
FT      cds
FT      56..2656
FT      Location/Qualifiers
FT      /tag= a
FT      /product=C-terminal alpha amidating enzyme
FT      WO9008190-A.
PN      26-JUL-1990.
PD      17-JAN-1990; J00042.
PR      17-JAN-1989; JP-005878.
PA      (SUNR ) SUNTORI LTD.
PI      Onsuye K, Kitano K, Tanaka S;
PI      WPI; 90-254034/33.
DR      P-PSDB; R96379.
PT      C-terminal alpha amidating enzymes from human thyroid gland -
PT      converting C-terminal glycyl peptide(s) or protein, and new DNA
PT      coding sequences.
PS      Disclosure; pp; English.
CC      CDNA libraries were prepared from human thyroid gland poly(A) RNA
CC      in lambda gtl1 and gtl0. These were screened using probes derived
CC      from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC      were isolated from the gtl0 library, including gtl0ch101.
CC      Restriction analysis indicated that gtl0ch101 encoded a different
CC      type of enzyme to one of the other two clones. The DNA can be
CC      inserted into vectors for expression in E.coli or (more efficiently)
CC      in animal cells.
CC      The enzymes are useful for prodn. of physiologically active alpha-
CC      amidated peptides from their C-glycyl precursors.
CC      See also Q05630 and Q05631.
CC      Sequence 3316 BP; 931 A; 676 C; 755 G; 954 T;

Query Match
Best Local Similarity 100.0%; Score 11; DB 1; Length 3316;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGTTCTAGT 11
DB      2813 CTGTTCTAGT 2823

RESULT 5
Q05630
ID      Q05630 standard; CDNA; 3688 BP.
AC      Q05630;
DE      17-DEC-1990 (first entry)
DE      Lambda gtl0ch201 encodes C-terminal alpha amidating enzyme.
DE      C-terminal alpha amidating enzyme; human thyroid gland;
KM      Lambda gtl0ch201; ss.
OS      Homo sapiens.
FH      Key
FT      cds
FT      140..3064
FT      Location/Qualifiers
FT      /tag= a
FT      /product=C-terminal alpha amidating enzyme.
FT      WO9008190-A.
PN      26-JUL-1990.
PD      17-JAN-1990; J00042.
PR      17-JAN-1989; JP-005878.
PA      (SUNR ) SUNTORI LTD.
PI      Onsuye K, Kitano K, Tanaka S;
PI      WPI; 90-254034/33.
DR      P-PSDB; R96399.
PT      C-terminal alpha amidating enzymes from human thyroid gland -
PT      converting C-terminal glycyl peptide(s) or protein, and new DNA
PT      coding sequences.
PS      Disclosure; pp; English.
CC      CDNA libraries were prepared from human thyroid gland poly(A) RNA
CC      in lambda gtl1 and gtl0. These were screened using probes derived
CC      from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC      were isolated from the gtl0 library, including gtl0ch201
CC      Restriction analysis indicated that gtl0ch201 encoded a different

```


CC type of enzyme to the other two clones. The DNA can be inserted
 CC into vectors for expression in E.coli or (more efficiently) in
 CC animal cells.
 CC The enzymes are useful for prodn. of physiologically active alpha-
 CC amidated peptides from their C-glycyl precursors.
 CC See also 005631 and 005632.
 SO Sequence 3688 BP; 1027 A; 775 C; 862 G; 1024 T;

Query Match 100.0%; Score 11; DB 1; Length 3688;
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 DB 3221 CTGTTCTAGT 3231

631
 005631 standard; DNA; 2863 BP.

AC Q05631;
 DE 17-DEC-1990 (first entry)
 DE Lambda gt10ch2 encodes C-terminal alpha amidating enzyme.
 KM C-terminal alpha amidating enzyme; human thyroid gland;
 KM Lambda gt10ch2; ss.
 OS Homo sapiens.
 FT key Location/Qualifiers
 FT cds 2..2332
 FT /tag= a
 FT /product=C-terminal alpha amidating enzyme

PN W09008190-A.
 PD 26-JUL-1990.
 PF 17-JAN-1990; J00042.
 PR 17-JAN-1989; JP-005878.
 PA (SUNR) SUNTORY LTD.
 PI Ohsuye K, Kitano K, Tanaka S;
 DR WPI: 90-234034/33.
 DR P-PSDB; R96400.
 PT C-terminal alpha amidating enzymes from human thyroid gland -
 PT converting C-terminal glycyl peptide(s) or protein, and new DNA
 PT coding sequences.
 PS Disclosure; pp: English.
 CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
 CC in Lambda gtl1 and gtl10. These were screened using probes derived
 CC from DNA encoding similar enzymes in Xenopus laevis. Three clones
 CC were isolated from the gtl10 library, including gt10ch2.
 CC Restriction analysis indicated that gt10ch2 encoded a different
 CC type of enzyme to one of the other two clones. The DNA can be
 CC inserted into vectors for expression in E.coli or (more efficiently)
 CC in animal cells.
 CC The enzymes are useful for prodn. of physiologically active alpha-
 CC amidated peptides from their C-glycyl precursors.
 CC See also 005630 and 005632.
 SO Sequence 2863 BP; 833 A; 572 C; 656 G; 800 T;

Query Match 100.0%; Score 11; DB 1; Length 2863;
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 DB 2489 CTGTTCTAGT 2499

RESULT 7
 ID Q11482
 AC Q11482;
 DE 19-JUN-1991 (first entry)
 DE Sequence encoding spike protein of infectious bronchitis virus.
 KM IBV; spike protein; vaccine; ss.

OS Infectious bronchitis virus.
 FH Key Location/Qualifiers
 FT cds 151..3657
 FT /tag= a
 FT /product= IBV spike protein
 FT 151..1767
 FT /tag= b
 FT /product= IBV spike protein S1 subunit
 FT 1783..3654
 FT /tag= c
 FT /product= IBV spike protein S2 subunit
 FT 1768..1782
 FT /tag= d
 FT /label= joining region

PN EP-423869-A.
 PD 20-APR-1991.
 PF 08-OCT-1990; 202667.
 PR 20-OCT-1989; US-424793.
 PA (AKRU) Akzo NV.
 PI Sondermeijer PJ, Claessens JA;
 DR WPI: 91-119142/17.
 DR P-PSDB; R11611.
 PT Nucleic acid sequence - encodes spike protein polypeptide of
 PT infectious bronchitis virus
 PS Disclosure; page 10; 20pp; English.
 CC The protein encoded by this sequence is a recombinant spike protein
 CC of infectious bronchitis virus (IBV) Arkansas strain. This protein
 CC or an antigenic fragment, e.g. S1 or S2 subunit, can be used in a
 CC vaccine for immunising poultry against IBV infections.
 CC See also Q11483-85.
 SO Sequence 3824 BP; 1114 A; 633 C; 733 G; 1344 T;

Query Match 100.0%; Score 11; DB 1; Length 3824;
 Best Local Similarity 100.0%; Pred. NO. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 DB 166 CTGTTCTAGT 176

RESULT 8
 ID Q13830
 AC Q13830 standard; DNA; 1920 BP.

DE 09-DEC-1991 (first entry)
 DE Msp gene in pUCRS (DSM 5803).
 KM Major secretion product; expression cassette; ss.
 OS Lactococcus lactis IM0230 (DSM 5805).
 FH Key Location/Qualifiers
 FT promoter 1..40
 FT /tag= a
 FT /note= "part of promoter region"

FT signal_peptide 411..491
 FT /tag= b
 FT mat_peptide 492..1793
 FT /tag= c

PN EP-449770-A.
 PD 02-OCT-1991.
 PF 13-MAR-1991; 810167.
 PR 22-MAR-1990; GB-006400.
 PA (CIBA) CIBA GEIGY AG.
 PI Surti B, Schmitz A;
 DR WPI: 91-250162/40.
 DR P-PSDB; R14150.
 PT Hybrid vectors for expression of polypeptide(s) - comprise DNA
 PT from plasmids obtd. from Lactococcus lactis, esp. major secretion
 PT prod. gene fragments.
 PS Disclosure; page 21; 36pp; English.
 CC The sequence (SDO ID NO:1) was obtd. from a genomic library of L.
 CC lactis in Lambda Emu1 3. The promoter region, and opt. the gene,
 CC can be fused with a gene encoding a protein for secretion of the

CC expressed protein into the supernatant.
 CC See also Q13831.
 SQ Sequence 1920 BP; 679 A; 344 C; 357 G; 540 T;

Query Match 100.0%; Score 11; DB 1; Length 1920;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 |||||
 DB 832 CTGTTCTAGT 842

RESULT 9
 Q14501
 ID Q14501 standard; DNA; 1592 BP.
 AC Q14501;
 PM 28-JAN-1992 (first entry)

USP45 gene.
 Chymosin; alpha-amylase; PN21011; enzyme; ss.
 OS Lactococcus lactis spp. lactis MG1363.
 FH Key Location/Qualifiers

FT cds 101..1483
 FT /*tag- a
 FT /label- dsp45
 FT signal-peptide 101..181
 FT /*tag- b
 FT mat-peptide 182..1483
 FT /*tag- c
 FT rbs 75..79
 FT /*tag- d
 FT rbs 90..93
 FT /*tag- e
 FT -10_signal 45..50
 FT /*tag- f
 FT -35_signal 25..30
 FT /*tag- g
 FT misc_signal 57
 FT /*tag- h
 FT /note- "start of transcription"
 FT repeat_unit 1514..1553
 FT /*tag- i
 FT /rpl_type- inverted
 FT /note- "rho-independent terminator"

PN NL9000753-A.
 16-OCT-1991.
 27-MAR-1991; 200721.
 30-MAR-1990; NL-000753.
 PA (NEZU-) NEDERLANDS INST ZUI.
 DR WPI; 91-330695/45.
 DR P-PSDB; R14530.

PT Lactococcus lactis DNA fragments - conty. extracellular protein
 PT signal peptide sequence
 PS Disclosure; Fig 4; 20pp; Dutch.
 CC The sequence is the KpnI/ClaI fragment from PN21011. Lactococci
 CC transformed with the gene may be used to produce extracellular
 CC heterologous enzymes of use in the food industry, e.g. chymosin or
 CC alpha-amylase.
 CC See also Q14502.

SO Sequence 1592 BP; 579 A; 291 C; 308 G; 414 T;

Query Match 100.0%; Score 11; DB 1; Length 1592;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 |||||
 DB 522 CTGTTCTAGT 532

RESULT 10

Q060820/C
 ID Q060820 standard; DNA; 391 BP.
 AC Q060820;
 DR 16-MAR-1994 (first entry)

DE Human brain Expressed Sequence Tag EST00924.
 KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 OS Homo sapiens.

PN WO9316178-A.
 PD 19-AUG-1993.
 PR 12-FEB-1993; U01294.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Adams MD, Moreno RF, Venter CJ;
 DR WPI; 93-272882/34.

PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 PS Example 4; Page 395-396; 500pp; English.
 CC The Expressed Sequence Tag was isolated from a human brain CDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prepn. of antisense sequences, probes and constructs.
 CC EST00924 has a "poor" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also Q59041-Q61440.
 SO Sequence 391 BP; 97 A; 106 C; 115 G; 71 T;

Query Match 100.0%; Score 11; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 |||||
 DB 76 CTGTTCTAGT 66

RESULT 11
 Q08935/C
 ID Q08935 standard; cDNA; 2232 BP.
 AC Q08935;
 DR 04-DEC-1995 (first entry)
 DE Human death associated protein DAP-1.
 KW Death associated protein; DAP; cytokine; cell death; ss.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT cds 160..468
 FT /*tag- a
 FT poly-a_signal 2203..2208
 FT /*tag- b
 FT misc_feature 160..468
 FT /*tag- c
 FT /note- "Claimed"
 FT misc_feature 287..818
 FT /*tag- d
 FT /note- "Claimed"

PN WO9510630-A.
 PD 20-APR-1995.
 PR 12-OCT-1994; U11598.
 PR 12-OCT-1993; IL-107250.

PA (RYCUC/) RYCUCS A.
 PI (YEDA) YEDA RES & DEV CO LTD.
 PI Kimchi A;
 DR WPI; 95-178528/23.
 DR P-PSDB; R74204.

PT DNA whose expression mediates cytokine-induced programmed cell
 PT death - used to treat diseases or disorders associated with
 PT uncontrolled, pathological cell growth or cytokine-induced
 PT programmed cell death.
 PS Claim 2; Fig 6; 61pp; English.

CC DAP genes seem to play an imp. role in programmed cell death and the
 CC inhibition of their expression protects the cell from cytokine-
 CC promoted cell death. An HL-60 cDNA library constructed in lambda
 CC 9t10 vector was screened with the cDNA insert of PTK01-230. Two
 CC independent clones, lambda1 and lambda2, almost completely
 CC overlapping and carrying cDNA inserts of about 2.3 kb were analysed.
 CC Lambda1 cDNA clone encompasses the 5'-UTR, short coding region(s)
 CC and a relatively long 3' UTR that constitutes more than 60% of the
 CC cDNA clone. The nt. sequence of the cDNA carried by lambda1 and
 CC predicted AA pattern are present in Q89835/R74204. The ORF is
 CC preceded by an extremely GC-rich 5' UTR and potentially codes for
 CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.
 CC It has an isoelectric pt. of 10 and is rich in prolines.
 CC Sequence 2232 BP; 542 A; 647 C; 569 G; 474 T;

Query Match 100.0%; Score 11; DB 1; Length 2232;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTGTTCTAGT 11
 |||||
 Db 636 CTGTTCTAGT 626

RESULT 12
 ID Q89837/c 089837 standard; cDNA; 531 BP.
 AC Q89837; standard; cDNA; 531 BP.
 DT 04-DEC-1995 (first entry)
 DE Human death associated protein DAP-1.
 KW Death associated protein; DAP; cytokine; cell death; ss.
 OS Homo sapiens.
 FH Key location/Qualifiers
 FT cds 1..182
 FT /tag- a
 FN W09510630-A.
 PD 20-APR-1995.
 PF 12-OCT-1994; U11598.
 PR 12-OCT-1993; IL-107250.
 PA (RVCU/) RVCUS A.
 PA (YEDA) YEDA RES & DEV CO LTD.
 PI Klmch1 A;
 DR WPI; 95-178528/23.
 PT DNA whose expression mediates cytokine-induced programmed cell
 death - used to treat diseases or disorders associated with
 uncontrolled, pathological cell growth or cytokine-induced
 programmed cell death.
 Claim 2; Fig 6; 61pp; English.
 CC DAP genes seem to play an imp. role in programmed cell death and the
 CC inhibition of their expression protects the cell from cytokine-
 CC promoted cell death. An HL-60 cDNA library constructed in lambda
 CC 9t10 vector was screened with the cDNA insert of PTK01-230. Two
 CC independent clones, lambda1 and lambda2, almost completely
 CC overlapping and carrying cDNA inserts of about 2.3 kb were analysed.
 CC Lambda1 cDNA clone encompasses the 5'-UTR, short coding region(s)
 CC and a relatively long 3' UTR that constitutes more than 60% of the
 CC cDNA clone. The nt. sequence of the cDNA carried by lambda1 and
 CC predicted AA pattern are present in Q89835/R74204. The ORF is
 CC preceded by an extremely GC-rich 5' UTR and potentially codes for
 CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.
 CC It has an isoelectric pt. of 10 and is rich in prolines. Q89837
 CC comprises a part of the ORF of Q89835 plus a part of the 3' UTR.
 CC Sequence 531 BP; 134 A; 153 C; 119 G; 125 T;

Query Match 100.0%; Score 11; DB 1; Length 531;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11
 |||||
 Db 349 CTGTTCTAGT 339

RESULT 13
 ID Q96296/c 096296 standard; DNA; 5836 BP.
 AC Q96296; standard; DNA; 5836 BP.
 DT 30-DEC-1995 (first entry)
 DE Human prostate-specific transcriptional regulatory element.
 KW Transcriptional regulatory element; prostate-specific antigen;
 KW prostate-specific enhancer; promoter; ss.
 OS Homo sapiens.
 FH Key location/Qualifiers
 FT enhancer 503..2974
 FT /tag- a
 FT /label- upstream PSA enhancer
 FT /note- "claimed, nts -5322 - -2851"

FT misc_signal 1..5824
 FT /tag- b
 FT /label- transcriptional regulatory element
 FT /note- "nts -5824 - -1"

FT misc_rna 5825..5831
 FT /tag- c
 FT /label- transcriptional start site
 FT /note- "nts +1 - +7"

FT promoter 5284..5831
 FT /tag- d
 FT /label- naturally-occurring PSA promoter
 FT /note- "nts -541 - +7"

FT promoter 5506..5831
 FT /tag- e
 FT /label- naturally-occurring PSA promoter
 FT /note- "nts -320 - +7"

FT misc_signal 524..3025
 FT /tag- f
 FT /label- transcriptional regulatory element
 FT /note- "claimed, nts -5300 - -2800"

FT enhancer 503..5824
 FT /tag- g
 FT /label- transcriptional regulatory element
 FT /note- "claimed, nts 0 - -5322"

PN W09519434-A.
 PD 20-JUL-1995.
 PF 12-JAN-1994; U00845.
 PR 13-JAN-1994; US-182247.
 PA (CALY-) CALYDON INC.
 PI Henderson DR.
 DR WPI; 95-263864/34.

PT Transcriptional regulator specific for cells expressing prostate specific
 PT antigen - used to express toxins, immunostimulants or anti-sense cpds.,
 PT for treatment and prevention of prostatic cancer or hypertrophy.
 PS Claim 2; Fig 1; 71pp; English.

CC Q96296 shows the sequence of the 5' flanking region of the human
 CC prostate specific antigen to -5824 bp. The fragment runs from a
 CC HindIII site at -5824 bp to the HindIII site at +7 bp. The
 CC numbering system is +1 at the transcriptional start site of PSA mRNA
 CC (Lundvall, A., 1989, Biochem. Biophys. Res. Comm. 161: 1151-59). The
 CC coding region of PSA starts at +42. An approx. 2.5 kbp fragment from
 CC about -2850 to about -5350 serves as a prostate specific enhancer
 CC (PSE) in conjunction with a promoter to function to initiate
 CC transcription in prostate cells. This can be isolated from the
 CC human genome as a XbaI-ApaI 2.5 kb fragment and this enhancer is
 CC termed the 'upstream PSA enhancer'. PSA - prostate-specific
 CC antigen. There is a naturally-occurring PSA promoter spanning
 CC the segment from about -541 to +7, particularly -320 to +7 of
 CC the human PSA gene. The promoter of PSA has been reported
 CC (Riegman et al., 1989, Biochem. Biophys. Res. Comm. 159: 95).
 CC The promoter from -320 to +7 contains a TATA-box, a GC-box and a
 CC hormonal response element at -170 to -156.
 CC Sequence 5836 BP; 1408 A; 1296 C; 1568 G; 1564 T;

Query Match 100.0%; Score 11; DB 1; Length 5836;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	Intron	1..55
FT	exon	/tag= a
FT	exon	56..155
FT	Intron	/tag= b
FT	exon	156..1512
FT	exon	/tag= c
FT	exon	1513..1611
FT	Intron	/tag= d
FT	Intron	1612..2206
FT	exon	/note= "n at 1925-1937 represent an indefinite interval within the Intron"
FT	exon	2207..2260
FT	Intron	/tag= f
FT	Intron	2261..2677
FT	exon	/tag= g
FT	exon	/note= "n at 2569-2581 represent an indefinite interval within the Intron"
FT	exon	2678..2788
FT	Intron	/tag= h
FT	Intron	2789..3328
FT	exon	/tag= i
FT	exon	/note= "n at 3063-3075 represent an indefinite interval within the Intron"
FT	exon	3329..3406
FT	Intron	/tag= j
FT	Intron	3407..3613
FT	exon	/tag= k
FT	exon	/note= "n at 3598-3610 represent an indefinite interval within the Intron"
FT	exon	3814..3902
FT	Intron	/tag= l
FT	Intron	3903..4224
FT	exon	/tag= m
FT	exon	/note= "n at 4076-4088 represent an indefinite interval within the Intron"
FT	exon	4225..4364
FT	Intron	/tag= n
FT	Intron	4365..6571
FT	exon	/tag= o
FT	exon	/note= "n at 4602-4614 represent an indefinite interval within the Intron"
FT	exon	6572..6677
FT	Intron	/tag= p
FT	Intron	6678..9163
FT	exon	/tag= q
FT	exon	9164..9307
FT	Intron	/tag= r
FT	Intron	9208..10530
FT	exon	/tag= s
FT	exon	10531..10607
FT	Intron	/tag= t
FT	Intron	10608..11597
FT	exon	/tag= u
FT	exon	/note= "n at 11383-11396 represent an indefinite interval within the Intron"
FT	exon	11598..15023
FT	Intron	/tag= v
FT	Intron	15024..15424
FT	exon	/tag= w
FT	exon	15425..15511
FT	Intron	/tag= x
FT	Intron	15512..15952
FT	exon	/tag= y
FT	exon	/note= "n at 15647-15659 represent an indefinite interval within the Intron"
FT	exon	15953..16126
FT	Intron	/tag= z
FT	Intron	16127..16565
FT	Intron	/tag= aa
FT	Intron	/note= "n at 16370-16382 represent an indefinite interval within the Intron"
FT	Intron	

FT	exon	16566..16692	/tag- ab	
FT	intron	16693..17535	/tag- ac	
FT	exon	/note- "n at 17290-17302 represent an indefinite interval within the Intron"		
FT	intron	17336..17726	/tag- ad	
FT	intron	17727..18416	/tag- ae	
FT	exon	/note- "n at 18299-18312 represent an indefinite interval within the Intron"		
FT	intron	18417..18787	/tag- af	
FT	exon	18788..19298	/tag- ag	
FT	intron	/note- "n at 18952-18964 represent an indefinite interval within the Intron"		
FT	exon	19299..19386	/tag- ah	
FT	intron	19387..20190	/tag- ai	
FT	exon	/note- "n at 19887-19899 represent an indefinite interval within the Intron"		
FT	intron	20191..20267	/tag- aj	
FT	intron	20268..21094	/tag- ak	
FT	exon	/note- "n at 20767-20779 represent an indefinite interval within the Intron"		
FT	intron	21095..21135	/tag- al	
FT	exon	21136..21583	/tag- am	
FT	intron	/note- "n at 21341-21353 represent an indefinite interval within the Intron"		
FT	exon	21584..21667	/tag- an	
FT	intron	21668..22233	/tag- ao	
FT	exon	/note- "n at 21921-21933 represent an indefinite interval within the Intron"		
FT	intron	22234..22288	/tag- ap	
FT	exon	22289..22832	/tag- aq	
FT	intron	/note- "n at 22567-22579 represent an indefinite interval within the Intron"		
FT	exon	22833..22906	/tag- ar	
FT	intron	22907..23287	/tag- as	
FT	exon	/note- "n at 23050-23062 represent an indefinite interval within the Intron"		
FT	intron	23288..23348	/tag- at	
FT	exon	23349..23698	/tag- au	
FT	intron	/note- "n at 23580-23592 represent an indefinite interval within the Intron"		
FT	exon	23699..24026	/tag- av	
FT	intron	2725	/tag- aw	
FT	exon	/note- "polymorphic site"		
FT	intron	3653	/tag- ax	
FT	exon	/note- "polymorphic site"		
FT	intron	4391	/tag- ay	
FT	exon	/note- "polymorphic site"		
FT	intron	4392	/tag- az	

FT		/note	"polymorphic site"
FT	misc_feature	/tag	6623
FT		/tag	ba
FT		/note	"polymorphic site"
FT	misc_feature	/tag	9106
FT		/tag	bb
FT		/note	"polymorphic site"
FT	misc_feature	/tag	9207
FT		/tag	bc
FT		/note	"polymorphic site"
FT	misc_feature	/tag	9376
FT		/tag	bd
FT		/note	"polymorphic site"
FT	misc_feature	/tag	11908
FT		/tag	be
FT		/note	"polymorphic site"
FT	misc_feature	/tag	11994
FT		/tag	bf
FT		/note	"polymorphic site"
FT	misc_feature	/tag	12952
FT		/tag	bg
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13004
FT		/tag	bh
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13009
FT		/tag	bi
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13048
FT		/tag	bj
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13238
FT		/tag	bk
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13448
FT		/tag	bl
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13539
FT		/tag	bm
FT		/note	"polymorphic site"
FT	misc_feature	/tag	13951
FT		/tag	bn
FT		/note	"polymorphic site"
FT	misc_feature	/tag	14041
FT		/tag	bo
FT		/note	"polymorphic site"
FT	misc_feature	/tag	14046

```

Query Match      100.0%; Score 11; DB 1; Length 24026;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 1 CTGTTCTAGT 11
    |||||
Db 13139 CTGTTCTAGT 13129

```

Search completed: September 18, 1999, 00:33:33
Job time: 18973 sec

OM nucleic - nucleic search, using sw model

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Searched: 2546578 seqs, 986266752 residues

:abase :

1:	em.est1.*
Est.*	
2:	em.est2.*
3:	em.est3.*
4:	em.est4.*
5:	em.est5.*
6:	em.est6.*
7:	em.est7.*
8:	em.est8.*
9:	em.est9.*
10:	em.est10.*
11:	em.est11.*
12:	em.est12.*
13:	em.est13.*
14:	em.est14.*
15:	em.est15.*
16:	em.est16.*
17:	em.est17.*
18:	em.est18.*
19:	em.est19.*
20:	qb.est1.*
21:	qb.est2.*
22:	qb.est3.*
23:	qb.est4.*
24:	qb.est5.*
25:	qb.est6.*
26:	qb.est7.*
27:	qb.est8.*
28:	qb.est9.*
29:	qb.est10.*
30:	qb.est11.*
31:	qb.est12.*
32:	qb.est13.*
33:	qb.est14.*
34:	qb.est15.*
35:	qb.est16.*
36:	qb.est17.*
37:	qb.est18.*
38:	qb.est19.*
39:	qb.est20.*
40:	qb.est21.*
41:	qb.est22.*
42:	qb.est23.*
43:	qb.est24.*
44:	qb.est25.*
45:	qb.est26.*
46:	qb.est27.*
47:	qb.est28.*
48:	qb.est29.*
49:	qb.est30.*
50:	qb.est31.*
51:	qb.est32.*
52:	em.est20.*
53:	em.est21.*

```
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB	ID
---------------	----------------	-----------------	----	----

Description

RESULT	1
L26727	
LOCUS	L26727 195 bp mRNA
DEFINITION	MUSEF00a. lambda untag male testis Mus musculus cDNA clone F300
ACCESSION	L26727
NID	9437510
VERSION	L26727.1 GI:437510

ALIGNMENTS

C	1	1	100.0	195	20	L26727	MUSF200A.1a
C	2	1	100.0	391	20	M8776	EST00924.Hu
C	3	1	100.0	369	20	T08787	EST06679.Hu
C	4	1	100.0	250	20	T10579	hbc233.Huma
C	5	1	100.0	183	20	T17272	N18615.Norm
C	6	1	100.0	234	20	T18606	hbc2266.Hu
C	7	1	100.0	265	20	T28324	EST32443.Hu
C	8	1	100.0	428	20	T36400	EST01330.S
C	9	1	100.0	346	20	T36417	EST01348.S
C	10	1	100.0	328	20	T48192	YD25H05.r1
C	11	1	100.0	294	20	T57378	YB51A04.r1
C	12	1	100.0	418	20	T57423	YB51A06.r1
C	13	1	100.0	513	20	T57520	YB58B01.s1
C	14	1	100.0	383	20	T235023	ATRS3662.Ac
C	15	1	100.0	295	20	T238765	HSCONB082.r
C	16	1	100.0	267	20	T238765	HSCOOD122.r
C	17	1	100.0	336	21	F01159	HSBA6E102.S
C	18	1	100.0	337	21	F04309	HSC2PD092.r
C	19	1	100.0	328	21	F04405	HSC2SB082.r
C	20	1	100.0	304	21	F04424	HSC2SFO32.r
C	21	1	100.0	280	21	F08472	HSCAVB061.r
C	22	1	100.0	358	21	F09540	HSC3A032.r
C	23	1	100.0	328	21	R04060	PK18B05.r1
C	24	1	100.0	392	21	R04483	PK24C04.r1
C	25	1	100.0	361	21	R06633	YF11B02.r1
C	26	1	100.0	342	21	R10668	YF31G10.s1
C	27	1	100.0	385	21	R12759	YF58H04.r1
C	28	1	100.0	436	21	R15422	YF48G08.r1
C	29	1	100.0	518	21	T65394	YF37C03.s1
C	30	1	100.0	394	21	T68242	YF40E09.r1
C	31	1	100.0	532	21	T77495	YF94G01.r1
C	32	1	100.0	539	21	T77506	YF94G12.r1
C	33	1	100.0	283	21	T87738	YF08D10.r1
C	34	1	100.0	343	22	H11107	Ym14D05.s1
C	35	1	100.0	431	22	R36022	YF69B06.r1
C	36	1	100.0	328	22	R45658	YF44E10.s1
C	37	1	100.0	398	22	R58601	YF92B01.s1
C	38	1	100.0	165	22	R57987	F7919.Feta
C	39	1	100.0	391	22	R66309	Y130G12.r1
C	40	1	100.0	431	22	R66315	Y130H12.r1
C	41	1	100.0	328	22	R70708	Y141H11.s1
C	42	1	100.0	326	22	R74865	MD80185R.Wc
C	43	1	100.0	234	23	D56671	HUM219C04B
C	44	1	100.0	248	23	D58198	HUM355A08B
C	45	1	100.0	439	54	HSW008808	A1043958.Homo.sapi


```

KEYWORDS      EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          1 (bases 1 to 195)
JOURNAL        Kerr, S.M., Vambrie, S., McKay, S.J. and Cooke, H.J.
MEDLINE        Analysis of cDNA sequences from mouse testis
COMMENT        Mamm. Genome 5, 557-565 (1994)
               On Apr 14, 1993 this sequence version replaced.

FEATURES
  source
    Location/Qualifiers
      1..195
        /organism="Mus musculus"
        /strain="Swiss"
        /db_xref="taxon:10090"
        /clone="F200"
        /note="Vector: lambda unizap male testis"
        /note="Vector: lambda unizap; lambda unizap library from
        mail mouse testis"

BASE COUNT    44 a      34 c      44 g      71 t      2 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 195;
Best Local Similarity 100.0%; Pred. No. 3e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 37 CTGTTCTACT 47

RESULT 2
LOCUS      M78776      391 bp      mRNA      EST      11-JAN-1995
DEFINITION EST00924 Hippocampus, strata gene (cat. #936205) Homo sapiens cDNA
clone HHCME73, mRNA sequence.
ACCESSION  M78776
NID        9273091
VERSION    M78776.1 GI:273091
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 391)
Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
Uterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
Sequence identification of 2,375 human brain genes
Nature 355, 632-634 (1992)
92168112
2 (bases 1 to 391)
Durbin, A.S.
EST mapping data
Unpublished (1996)

COMMENT
Contact: Kerlavage, AR
Bioinformatics for Genomic Research
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.

FEATURES
  source
    Location/Qualifiers
      1..391
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):78208"
        /db_xref="GDB:D051760E"
        /db_xref="taxon:9606"

```

```

/map="22"
/clone="HHCME73"
/clone_lib="Hippocampus, Strata gene (cat. #936205)"
/note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +
random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
average insert size."

BASE COUNT    97 a      106 c      115 g      71 t      2 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 266 CTGTTCTACT 276

RESULT 4
LOCUS      T10579      250 bp      mRNA      EST      29-NOV-1993
DEFINITION hbc233 Human pancreatic islet Homo sapiens cDNA clone hbc233 5'end,
mRNA sequence.
ACCESSION  T10579
NID        9390733

FEATURES
  source
    Location/Qualifiers
      1..369
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):85285"
        /db_xref="taxon:9606"
        /clone="H18J71"
        /clone_lib="Infant Brain, Bento Soares"

BASE COUNT    60 a      119 c      97 g      92 t      1 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 266 CTGTTCTACT 276

RESULT 3
LOCUS      T08787      369 bp      mRNA      EST      03-AUG-1993
DEFINITION EST06679 Infant Brain, Bento Soares Homo sapiens cDNA clone H18J71
5'end, mRNA sequence.
ACCESSION  T08787
NID        9389815
VERSION    T08787.1 GI:389815
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 369)
Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
Nature Genet. 4, 373-380 (1993)
94004965

COMMENT
Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Galthersburg, MD 20878
Tel: 3018699036
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13 Reverse.

FEATURES
  source
    Location/Qualifiers
      1..369
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):85285"
        /db_xref="taxon:9606"
        /clone="H18J71"
        /clone_lib="Infant Brain, Bento Soares"

BASE COUNT    60 a      119 c      97 g      92 t      1 others
ORIGIN
Query Match   100.0%; Score 11; DB 20; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTACT 11
Db 266 CTGTTCTACT 276

RESULT 4
LOCUS      T10579      250 bp      mRNA      EST      29-NOV-1993
DEFINITION hbc233 Human pancreatic islet Homo sapiens cDNA clone hbc233 5'end,
mRNA sequence.
ACCESSION  T10579
NID        9390733

```


VERSION T10579.1 GI:390733
 EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 250)
 AUTHORS Takeda, J., Yano, H., Eng, S., Zeng, Y. and Bell, G. I.
 TITLE A molecular inventory of human pancreatic islets: sequence analysis
 of 1000 cDNA clones
 JOURNAL Hum. Mol. Genet. 2, 1793-1798 (1993)
 MEDLINE 94108427
 COMMENT Contact: Bell GI or Takeda J
 HHMI
 Univ. of Chicago
 5841 S. Maryland Ave., MC1028, Chicago IL 60637
 Tel: 3127029116
 Fax: 3127020271
 Email: g-bell@uchicago.edu
 Seq primer: SK primer.
 Location/Qualifiers
 1..250
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbc233"
 /note="Vector: Lambda ZAPII; Site 1: Eco RI; Site 2: Xho
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."
 BASE COUNT 60 a 47 c 63 g 80 t
 ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 250;
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 |||||
 Db 82 CTGTTCTAGT 92

RESULT 5
 --7272/c
 US T17272 183 bp mRNA EST 25-JUL-1996
 DEFINITION NTB615 Normalized infant brain, Bento Soares Homo sapiens CDNA
 3' end, mRNA sequence.
 ACCESSION T17272
 NID 9519434
 VERSION T17272.1 GI:519434
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 183)
 AUTHORS Berry, R., Stevens, T. J., Walter, N. A. R., Wilcox, A. S., Rubano, T.,
 Hopkins, J. A., Weber, J., Gould, R., Soares, M. B. and Sikele, J. M.
 TITLE Gene-based Sequence Tagged Sites (STSS) as the basis for a human
 gene map
 JOURNAL Nature Genet. 10, 415-423 (1995)
 MEDLINE 95400322
 COMMENT Contact: Sikele JM
 Department of Pharmacology
 University of Colorado Health Sciences Center
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
 Tel: 3032708637
 Fax: 3032707097
 Email: nikki@tally.uchsc.edu

FEATURES Seq primer: -21M13 Universal.
 Location/Qualifiers
 1..183
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbc2266"
 /note="Vector: Lambda ZAPII; Site 1: Eco RI; Site 2: Xho
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."
 BASE COUNT 79 a 46 c 40 g 69 t
 ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 234;
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGTTCTAGT 11
 |||||
 Db 44 CTGTTCTAGT 34

RESULT 6
 LOCUS T18606 234 bp mRNA EST 28-APR-1994
 DEFINITION hbc2266 Human pancreatic islet Homo sapiens CDNA clone hbc2266 5',
 mRNA sequence.
 ACCESSION T18606
 NID 9475414
 VERSION T18606.1 GI:475414
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 234)
 AUTHORS Bell, G. I. and Takeda, J.
 TITLE Human pancreatic islet cDNAs
 JOURNAL Unpublished (1995)
 COMMENT On Sep 21, 1992 this sequence version replaced gi:276118.
 Contact: Bell GI or Takeda J
 HHMI
 Univ. of Chicago
 5841 S. Maryland Ave., MC1028, Chicago IL 60637
 Tel: 3127029116
 Fax: 3127020271
 Email: g-bell@uchicago.edu
 Seq primer: SK primer.
 Location/Qualifiers
 1..234
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbc2266"
 /note="Vector: Lambda ZAPII; Site 1: Eco RI; Site 2: Xho
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."

QY 1 CTTTCTAGT 11
|||||

Db 82 CTTTCTAGT 92

RESULT 7
T28224 265 bp mRNA EST 06-SEP-1995
LOCUS T28224
DEFINITION EST32443 Human Embryo Homo sapiens CDNA 5' end similar to
peptidylglycine alpha-amidating monooxygenase (HT:1157), mRNA
sequence.
ACCESSION T28224
NID 9610322
VERSION T28224.1 GI:610322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 265)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fudner,R.A.,
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georghagen,N.S.M.,
Glodek,A., Gnehm,C.U., Hanna,M.C., Hedblom,E., Hinkley,F.S.,
Kelley,J.M., Kline,K.M., Kelley,J.C., Liu,L.-I., Matmaros,S.M.,
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,
Pellegriano,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R.,
Weidman,T.F., Li,Y., Bednarek,D.P., Ceol,L., Cepeda,M.A.,
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrite,A.,
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
Gibber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Li,H., Li,H.,
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
Haseeltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns
Based Upon 83 Million Basepairs of CDNA Sequence
Nature 377, 3-174 (1995)
96026280

TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Venter, J.C.
The Institute for Genomic Research
932 Clopper Rd, Gathersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1..265
/organism="Homo sapiens"
/db_xref="ATCC (inhost):101601"
/db_xref="taxon:9606"
/clone_id="Human Embryo"
/tissue_type="Embryo"
BASE COUNT 68 a 52 c 43 g 101 t 1 others
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 265;
Best Local Similarity 100.0%; Pred. No. 2.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTCTAGT 11
|||||

Db 52 CTTTCTAGT 62

RESULT 8

T36400 428 bp mRNA EST 11-JAN-1995
LOCUS T36400
DEFINITION EST101330 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
CDNA 3' end, mRNA sequence.
ACCESSION T36400
NID 9620217
VERSION T36400.1 GI:620217
KEYWORDS EST.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 428)
Weinstock,K.
Saccharomyces cerevisiae cDNAs
Unpublished (1995)
Other ESTs: TC2
Contact: Weinstock,K. and Venter,J.C.
The Institute for Genomic Research
932 Clopper Rd, Gathersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13-21.

FEATURES
Location/Qualifiers
1..428
/organism="Saccharomyces cerevisiae"
/strain="X2180-1A"
/db_xref="taxon:4932"
/clone_id="S. cerevisiae strain X2180-1A"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 106 a 93 c 72 g 157 t
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 428;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTCTAGT 11
|||||

Db 119 CTTTCTAGT 129

RESULT 9

T36417 346 bp mRNA EST 11-JAN-1995
LOCUS T36417
DEFINITION EST101348 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
CDNA 3' end, mRNA sequence.
ACCESSION T36417
NID 9620234
VERSION T36417.1 GI:620234
KEYWORDS EST.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 346)
Weinstock,K.
Saccharomyces cerevisiae cDNAs
Unpublished (1995)
Other ESTs: TC2
Contact: Weinstock,K. and Venter,J.C.
The Institute for Genomic Research
932 Clopper Rd, Gathersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@tdb.tigr.org)

FEATURES
source
Seq primer: M13-21.
Location/Qualifiers
1. .346
/organism="Saccharomyces cerevisiae"
/strain="X2180-1A"
/db_xref="taxon:4932"
/clone_1lb="S. cerevisiae strain X2180-1A"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT
81 a 68 c 63 g 130 t 4 others

ORIGIN
1 CTGTTCTAGT 11
|||||
119 CTGTTCTAGT 129

Query Match
Best Local Similarity 100.0%; Score 11; DB 20; Length 346;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
T48192/c 328 bp mRNA EST 01-FEB-1995
LOCUS yb51805.r1 StrataGene fetal spleen (#937205) Homo sapiens cDNA
DEFINITION
clone IMAGE:7249 5', mRNA sequence.
ACCESSION
T48192
NID 9650172
VERSION T48192.1 GI:650172
KEYWORDS
EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Roifling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 221
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RP1
High quality sequence stop: 221.
Location/Qualifiers
1. .328
/organism="Homo sapiens"
/db_xref="GDB:493914"
/db_xref="taxon:9606"
/map="X; 2: 726H05: 12"
/clone="IMAGE:72249"
/clone_1lb="StrataGene fetal spleen (#937205)"
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: spleen; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 Kb;
Oligo dt. Pooled spleens. Average insert size: 1.0 Kb;

FEATURES
source
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGACGACG
3' ~3' adaptor sequence: 5' CTCAGATTTTCTTTTCTTTT 3'

BASE COUNT
71 a 77 c 103 g 71 t 6 others

ORIGIN
1 CTGTTCTAGT 11
|||||
182 CTGTTCTAGT 172

Query Match
Best Local Similarity 100.0%; Score 11; DB 20; Length 328;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
T57378 294 bp mRNA EST 08-FEB-1995
LOCUS yb51804.r1 StrataGene fetal spleen (#937205) Homo sapiens cDNA
DEFINITION
clone IMAGE:74670 5', mRNA sequence.
ACCESSION
T57378
NID 9659239
VERSION T57378.1 GI:659239
KEYWORDS
EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Roifling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 704
High quality sequence stops: 216 Source: IMAGE Consortium, LNL This
clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 704 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 216.
Location/Qualifiers
1. .294
/organism="Homo sapiens"
/db_xref="GDB:496335"
/db_xref="taxon:9606"
/clone="IMAGE:74670"
/clone_1lb="StrataGene fetal spleen (#937205)"
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: spleen; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 Kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGACGACG
3' ~3' adaptor sequence: 5' CTCAGATTTTCTTTTCTTTT 3'";

BASE COUNT
79 a 65 c 60 g 84 t 6 others

ORIGIN
100.0%; Score 11; DB 20; Length 294;
Query Match

Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTTCTAGT 11
|||||
Db 200 CTGTTCTAGT 210

RESULT 12

TS7423 418 bp mRNA EST 08-FEB-1995
LOCUS TS7423/c
DEFINITION yb58b01.s1 Stratagene fetal spleen (#937205) Homo sapiens CDNA
clone IMAGE:74722.5, similar to similar to gb:U04217
PORPHOBILINOGEN DEAMINASE (HUMAN), mRNA sequence.

ACCESSION TS7423
NID 9659284
VERSION TS7423.1 GI:659284
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chappell, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1707
High quality sequence stops: 291 Source: IMAGE Consortium, LNL. This
clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 1707 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 291.
Location/Qualifiers

FEATURES
source 1. 418
/organism="Homo sapiens"
/db_xref="GDB:496387"
/db_xref="taxon:9606"
/clone="IMAGE:74722"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: spleen; Vector: Bluescript SK-; site_1:
EcoRI; site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Pooled spleens. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'";
BASE COUNT 98 a 95 c 112 g 105 t 8 others
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 418;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGTTCTAGT 11
|||||
Db 205 CTGTTCTAGT 195

RESULT 13

TS7520 513 bp mRNA EST 08-FEB-1995
LOCUS TS7520
DEFINITION yb58b01.s1 Stratagene ovary (#937217) Homo sapiens CDNA clone
IMAGE:75337.3, similar to contains Alu repetitive element, mRNA
sequence.

ACCESSION TS7520
NID 9659381
VERSION TS7520.1 GI:659381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chappell, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1280
High quality sequence stops: 321 Source: IMAGE Consortium, LNL. This
clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information. Possible
reversed clone: polyT not found
Insert Length: 1280 Std Error: 0.00
Seq primer: -21m13
High quality sequence stop: 321.
Location/Qualifiers

FEATURES
source 1. 513
/organism="Homo sapiens"
/db_xref="GDB:505066"
/db_xref="taxon:9606"
/clone="IMAGE:75337"
/clone_lib="Stratagene ovary (#937217)"
/sex="female"
/dev_stage="49 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ovary; Vector: Bluescript SK-; site_1:
EcoRI; site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Total ovary tissue, normal, caucasian. Average insert
size: 0.8 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'
GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'";
BASE COUNT 114 a 104 c 112 g 180 t 3 others
ORIGIN

Query Match 100.0%; Score 11; DB 20; Length 513;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGTTCTAGT 11
|||||
Db 123 CTGTTCTAGT 133

RESULT 14
235023/c

LOCUS 235023 383 bp mRNA EST 18-JUL-1994
 DEFINITION AT35362 AC16H Arabidopsis thaliana cDNA clone TA1082.3', mRNA
 sequence.
 ACCESSION 235023
 MID 9510771
 VERSION 235023.1 GI:510771
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euophylliophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 REFERENCE 1 (bases 1 to 383)
 CNRS.
 The Arabidopsis thaliana transcribed genome: the GDR cDNA program
 unpublished (1996)
 JOURNAL COMMENT
 CONTACT: Berthomieu P., Guerrier D., Giraudat J.
 Genetique Molculaire d'Arabidopsis
 ISV - UPR40, CNRS
 Avenue de la Terrasse, 1198 Gif-sur-Yvette Cedex, France
 Email: Giraudat@cnrs-gif.fr.
 FEATURES
 SOURCE
 1. 383
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="ecotype Columbia"
 /db_xref="taxon:3702"
 /clone="TA1082"
 /clone_1lb="AC16H"
 /note="Vector: Lambda ZAPIT; tissue-cell suspension
 culture of ecotype Columbia; clone_library=AC16H; Cloning
 vector: Lambda ZAPIT; Physiological condition: cycling
 cells."
 BASE COUNT 106 a 82 c 69 g 121 t 5 others
 ORIGIN
 Query Match 100.0%; Score 11; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGTTTCAGT 11
 |||||||||
 Db 129 CTGTTTCAGT 119
 |||||||||
 U1T 15
 /728
 LOCUS 238728 295 bp mRNA EST 31-OCT-1994
 DEFINITION HSCONB082 normalized infant brain cDNA Homo sapiens cDNA clone
 c-0nb08.3', mRNA sequence.
 ACCESSION 238728
 MID 9561120
 VERSION 238728.1 GI:561120
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 295)
 Aulifrey, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
 Devignes, M.D., Duprat, S., Houligatte, R., Juneau, M.N., Lamy, B.,
 Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,
 Sebastiani-Kabackich, C. and Tessier, A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 JOURNAL MEDLINE
 COMMENT 95277534
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read. 18 T removed at sequence 5' end
 Genexpress library_id: C; Genexpress_sequence_id: alc-0nb08
 Seq primer: (-21)M3-universal.
 FEATURES
 SOURCE
 1. 295
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="c-0nb08"
 /clone_1lb="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lambda BA; Site_1: HindIII;
 Site_2: NotI; sex=Female; dev_stage=3 months old;
 isolate=muscular atrophy patient; tissue_type=total
 brain; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambda BA vector. Clone library from B. Soares, Psychiatry
 Dept. Columbia University, USA
 Bento Soares, P.N.A.S. in press" Normalization_method:
 BASE COUNT 81 a 60 c 62 g 92 t
 ORIGIN
 Query Match 100.0%; Score 11; DB 20; Length 295;
 Best Local Similarity 100.0%; Pred. No. 2.8e+03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGTTTCAGT 11
 |||||||||
 Db 23 CTGTTTCAGT 33
 |||||||||

Search completed: September 17, 1999, 21:28:00
 Job time: 14264 sec


```

/db_xref="PID:92622144"
/db_xref="GI:2622144"
/translation="MSTIPGOLFEPDSSGDKITAKSTSLPKVDAAKKYKKYKRYK
HASKGYAKKIRYKTKYRTAAKTKRKRYRSATYTSRKAYRAISDDL
NNLOGDEBELKASYINNRLNHRSGPETAAGVGTGDCWGLSDMSARLANAGK
VKVVOGATSSSSRRHMLHVYSEGRWTSPELVTMKRYGSKRYTATCGRAFRVATYNN
"
4683..5372
/gene="MTH1041"
/feature="CDS"
4683..5372
/gene="MTH1041"
/note="Function Code:1.01 - Carbohydrate Metabolism,
glycolysis--gluconeogenesis ; similar to, sp:LN:RITS_METRA
AC:058923, p(1)-1E-63, pld=51%"
/codon_start=1
/transl_table=11
/product="triosephosphate isomerase"
/protein_id="AAB85537.1"
/db_xref="PID:92622145"
/db_xref="GI:2622145"
/translation="MLEDLEKDTPIYITLNFKYLESTGERALELASICGDVADETV
NMAYAFQMDLHRVSDAVEIPVLAQHIDAVDAGHTGTSIIAECARDAGAAGTIIHSE
KRMQLADIEMWISMKLEEMSVVCTNNMTTAAALGDFEVAVEPELIGGIPYS
RAPEEVTIGSDAVKKNPEVSYLCGAGISTGDDMKAAVLDGAEGVLLAGCITLADSP
RDALDLVSKY"
5369..6604
/gene="MTH1042"
/feature="CDS"
5369..6604
/gene="MTH1042"
/note="Function Code:1.01 - Carbohydrate Metabolism,
glycolysis--gluconeogenesis ; similar to, sp:LN:PEK_METRE
AC:820971, p(1)-2.3E-137, pld=62%"
/codon_start=1
/transl_table=11
/product="3-phosphoglycerate kinase"
/protein_id="AAB85538.1"
/db_xref="PID:92622146"
/db_xref="GI:2622146"
/translation="MMYIPVSFKRTMDIEVTKYLVANVINSVPDNDGILTDT
RMLHATTIRELSDRGARTVMAHQSPGKNDPTTLEQHRALSGILRPVKVEDIF
GSAAREKISGLGGEITLLENVRFSESEVLRKDEPDAETHLVKLAFLIDYFINAF
AAAHRSQPSLVGFALRVPSAAGVMESELTLOGALENRPVYVIGGVKVDISNV
MKNVLENGSADLVLTGLVANIFLAGGVVIGKVMDEFIKSRGCDFFIVAKLKRFR
PERIIVPVDAVCRDGRVDVPVKIPIPNHIODIGMETITLYARRIRERTLEKRP
GVFENPESIGTEIDINAISSSGFSLIGGHLAAAVAKMGFEEDSINHISGGGASIS
LAGEELPRVRYLEESNHP"
Complement(6645..6718)
/note="74 bp inverted repeat includes part of MTH1043
(trNA-Thr anticodon:ugu); 84% ID to interval
678954-679026"
6645..6718
/gene="MTH1043"
/note="codon recognized: ACA"
/product="trNA-Thr"
6645..6718
/gene="MTH1043"
/complement(6651..6711)
/note="61 bp direct repeat includes part of MTH1043
(trNA-thr anticodon:ugu); 85% ID to interval
954496-954436"
6726..6800
/gene="MTH1044"
/note="codon recognized: CCA"
/product="trNA-Pro"
6726..6800
/gene="MTH1044"
/feature="CDS"
6835..6927
/gene="MTH1045"
/note="codon recognized: TAC"
/product="trNA-Tyr"
6835..6927
/gene="MTH1045"
/feature="CDS"
7056..7127

```

```

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 18988;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 2660 ACCCACACAG 2651

RESULT 3
ANNOXEF/c ANNOXEF 2280 bp DNA BCT 01-DEC-1998
LOCUS Anacystis nidulans hoxe, ORF1 genes.
DEFINITION Y13471
ACCESSION Y13471
VERSION 93947769
KEYWORDS hoxe gene; hoxf gene; ORF1.
SOURCE Synecchococcus PCC6301.
ORGANISM Synecchococcus PCC6301
Bacteria: Cyanobacteria; Chroococcales; Synecchococcus.
REFERENCE 1 (bases 1 to 2280)
AUTHORS Boisson,G., Schmitz,O., Schmitz,B. and Bothe,H.
TITLE Unusual gene arrangement of the bidirectional hydrogenase and
functional analysis of its diaphorase subunit Hoxf in respiration
of the unicellular cyanobacterium anacystis nidulans
JOURNAL Curr. Microbiol. 36 (5), 253-258 (1998)
MEDLINE 98210061
REFERENCE 2 (bases 1 to 2280)
AUTHORS Boisson,G.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1997) G. Boisson, Universitaet zu Koeln,
Botanisches Institut, Gyrhofstr. 15, D-50931 Koeln, FRG
FEATURES
Source Location/Qualifiers
1..2280
/organism="Synecchococcus PCC6301"
/strain="1402-1"
/db_xref="taxon:1139"
/clone_lib="lambdaGEM-11"
53..>2280
/gene="hoxe"
/note="putative"
53..2280
/gene="hoxe"
74..571
/gene="hoxe"
/codon_start=1
/transl_table=11

```


gene
RBS
CDS

559..2175
/gene="hoxe"
559..563
/gene="hoxe"
/note="putative"
571..2175
/gene="hoxe"
/codon_start=1
/transl_table=1
/protein_id="CA73873.1"
/db_xref="PID:e321437"
/db_xref="PID:g3947771"
/db_xref="GI:3947771"
/translation="MDWEDLGRLANEELTCQPIRLRCCATGCRANGAEAVFRAVOO
LTPAONLGRCEAVSGLGCGAGPLVOCQPSRLYSDIPDOADLVAAQGAAND
LPEVDOAPFERSQOLKIVNRHSGILNDRLESYLAAGYRLMTITDITPTVEVETI
RLSLRGRGGGGPTGLKMATYAKKPSDRKRVVNCGBDGPDAINDRSYLSDFHOYT
EGMIAIYAVGANFGYLVRAEYPLAIRLOAIRQARRGLGNSYLDSEFSFDELE
RIGAFVYCEGETALIHISIOGERVPRVPRYPAPESGLMGPTLINNVEPTANAPITV
EQGDWFAIETPTSKGTGKVFALTKLNGLIEVPMGIPRSIVDGIPESEVKAV
OTGSPSGGCIPLAODPVDPSLIOLGSMXSGGMVMDENTMDVAIARYMEFERS
ESGCKCIPCRAGTYQVHLHELKSSGGGTAIDLOQLEDLCIVADTSLCIGMSAPNPN
ILSLTRNFRQETESRLPERAIALTH"
complement(2185..>2280)

CDS

BASE COUNT 475 a 643 c 670 g 492 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 2280;
Jacc Local Similarity 100.0%; Prid. No. 3.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 463 ACCCACACAG 454

1 ACCCACACAG 10
|||||||

RESULT 4
AP000002 BCT 08-FEB-1999
LOCUS Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position
DEFINITION (277)
ACCESSION AB0009475 AB009476 AB009477 AB009478 AB009479 AB009480
AB009481 AB009482 AB009483 AB009484
NID 93236129
KEYWORDS AP000002.1 GI:3236129

SOURCE
ORGANISM Pyrococcus horikoshii (strain:OT3) DNA.
Pyrococcus horikoshii
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.

REFERENCE 1 (bases 1 to 257000)
AUTHORS Tanaka,T., Kawarabayashi,Y. and Kikuchi,H.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Kawarabayashi, National Institute of Technology and Evaluation,
Biotechnology Center; 2chome 49-10 Nishihara, Shibuya-W., Tokyo
151-0066, Japan (E-mail: genome@nitech.go.jp, Tel:+81-3-3481-8951,
Fax:+81-3-3481-8424)
2 (sites)

Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y.,
Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,
Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohnuku, Y.,
Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N.,
Oyuchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K.,
Masuch, Y., Shikuya, H. and Kikuchi, H.
Complete sequence and gene organization of the genome of a
hyperthermophilic archaeobacterium, Pyrococcus horikoshii OT3
DNA Res. 5 (2), 55-76 (1998)
98344137

On Jul 28, 1998 this sequence version replaced g1:3130553
g1:3130577 g1:3130627 g1:3130656 g1:3130685 g1:3130705 g1:3130737
g1:3130785 g1:3130824 g1:3130828
AB009475-AB009484: submitted (10-DEC-1997)
Kawarabayashi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuch, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.

Shizuya, H. is at the California Institute of Technology, Pasadena,
CA, USA.
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS

Homology analysis is performed by Smith-Waterman algorithm against
GenBank and Genpept release 103; EMBL release 52.0; SwissProt
release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genome@nitech.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: http://www.blo.nite.go.jp/.

Location/Qualifiers

1. .257000
/organism="Pyrococcus horikoshii"
/strain="OT3"
/db_xref="taxon:53953"
/map="287001-54400 nt. position"
1363. .2664
/gene="PH0318"
1363. .2664
/gene="PH0318"
/note="similar to PIR:B64485 percent identity: 55.181 in
403aa; PIR:S74862 percent identity: 49.869 in 397aa;
Swiss_Prot:P33221 percent identity: 50.773 in 404aa"
/codon_start=1
/transl_table=11
/product="433aa long hypothetical
phosphoribosylglycinamide formyl transferase"
/protein_id="BA29382.1"
/db_xref="PID:d103035"
/db_xref="PID:g3256709"
/db_xref="GI:3256709"
/translation="NYVMIKRLDELGTATSDSAOKILLSGELGEKEIAIEAORLGE
VAVADVAPNAPAMOVAAHRSRVGVMMDMKDFMSVSEPEKPAIIPETAEINLDLPFE
KDQFPVPNARATIMMRNRLEKETLYREKAYPTSRMYATTIDELTEACEKIGTGYCH
TKAIMSSGSGSTFVGPEPDIPIAMEDEAKRAGSAEKILVEEHIDFDEVVELAVRH
FDNGEIVTFEPKPVGHYQLDGDYHAASWOPAESKAERVRYPAIARIIDVDLGGLIF
GVMEFGKGRDVANVESPRHDHGVTGLASHPGEGSEFLALRAVLGDLPIGEMVDGY
RLPEMLIPATVAHVIAKAGSGSPRGVLVALSVPNATVATFKCPFAVYGRRLGIALA
WDVDVAKRKAEVVAHMTELRTRSSMDHQNEKRKHLLR"
complement(1645..1974)
/gene="PH0319"
complement(1645..1974)

FEATURES
source

gene
CDS

gene
CDS

CDS


```

/gene="PH0319"
/codon_start=1
/transl_table=11
/product="109aa long hypothetical protein"
/protein_id="BAA29393.1"
/db_xref="PID:d1030336"
/db_xref="PID:g3256710"
/db_xref="GI:3256710"
/transl_table="MFLNYLCTSFSCFPFGRNVLGPDEVRSLRGHYSIC
VAVITDFTGLIEFKGCSVHIPGSRNLCLHKGFPKPLPVHSPGSSCIWYKITYL
LEFEEG"
/complement(2653. .3186)
/gene="PH0320"
/complement(2653. .3186)
/gene="PH0320"
/notes="similar to owl:SSOURK percent identity: 68.000 in
150aa; PIR:S76001 percent identity: 60.694 in 173aa;
owl:BAPURKE1 percent identity: 62.329 in 146aa"
/codon_start=1
/transl_table=11
/product="177aa long hypothetical
phosphoribosylaminoimidazole carboxylase catalytic
subunit"
/protein_id="BAA29394.1"
/db_xref="PID:d1030337"
/db_xref="PID:g3256711"
/db_xref="GI:3256711"
/transl_table="MVISMKSEKPLVGIINGSDPLPVMKAARIIEFQVPEITII
SAHTRPRAEYAKAEERGIIEVLIAGGAHLPGLIASLTVLPYIGVIEKALNG
LDSLSITVOMPSGIPVATVAIDNAKNALLALNLGIKYEIEKRLRRYKDKRRVE
EKAKRLEEMGEMERTLSE"
/complement(3275. .4036)
/gene="PH0321"
/complement(3275. .4036)
/gene="PH0321"
/codon_start=1
/transl_table=11
/product="253aa long hypothetical protein"
/protein_id="BAA29395.1"
/db_xref="PID:d1030338"
/db_xref="PID:g3256712"
/db_xref="GI:3256712"
/transl_table="MISWEIEELDEIRGKIKKSLIIIEEDSSSGKDLFFILSRK
LKSNDLVGMFSISPIOLITIRISREVDVTKYLENRLIVTFISFQIKTMGV
WYLEGMLSSSETLPKIKYAKAVEDHKKVMMDLNEGRLYGFAISMGVLEFPEETL
RYLETSAEVRGHPAYKKYRGNTFMMEGVKRRVLLSVYRRADYVLKTRSSILGNG
IKRELIVIKTPKPIEELVREYEFKGNPKLRVD"
/complement(4039. .4698)
/gene="PH0322"
/complement(4039. .4698)
/gene="PH0322"
/codon_start=1
/transl_table=11
/product="219aa long hypothetical protein"
/protein_id="BAA29396.1"
/db_xref="PID:d1030339"
/db_xref="PID:g3256713"
/db_xref="GI:3256713"
/transl_table="MGIIEELVSIPEGGILATIIQKELESGDRFGILVYLKLENN
NVLFLTEPELRTMNTMENGIVNDEIGENMLIPVFGSMNIEKIKGVYLSYL
DDYVYSKLKEMAKIKTKNTINANFEFTYLSGGICLISNPLTKLITKLELKEILK
GYRKITITISQLECPILIEETVLAISVLETRIINGKKGVIITKGNENLIELEKE
V"
/4856. .6172
/gene="PH0323"
/4856. .6172
/gene="PH0323"
/notes="similar to PIR:A64417 percent identity: 55.478 in
442aa; owl:HUMRPS4 percent identity: 33.738 in 444aa;
Swiss_Prot:P12039 percent identity: 33.819 in 352aa.
motif-prokaryotic membrane lipoprotein lipid attachment
site"
/codon_start=1

```

```

/transl_table=11
/product="438aa long hypothetical
phosphoribosylamide--glycine ligase"
/protein_id="BAA29397.1"
/db_xref="PID:d1030340"
/db_xref="PID:g3256714"
/db_xref="GI:3256714"
/transl_table="MKVLIVGGGGRHAIGELVKGAEIYVSNHRNPGMLRKADY
GLAKETNVEDVTKFARKWGIELAFIGEPNPLAGVYNALDEEGIPAVGPTREARLET
NKAMAREEPERNINPGKMFRIEDVQEMRKNIDEXGKVVVKPLGLTGKGVKVVGY
OLKDNDEAKAYAEHIIKDKGVLIBERTDGVFTLOFTDGVKVPMPVQYPAHYE
GVGPRTGSGVSCSNHLLPRTGDFERALKTEETEARKRGVPEPKGLYQFM
LSGEPVLEIYNARBDPEAINVLAIVLDNPLIEIKAGIYEGSRKAKELNKATVYKI
APOGIPQDPIKGIKRIEVDDEGILNKGAKITIVAVDENLILGSRALATVGVADSLSEA
ERIAENGVSIVKGPPIFYRNDVGTRESVERKRIEMKKGKEFEPNLC"
/6174. .7325
/gene="PH0324"
/6174. .7325
/gene="PH0324"
/notes="similar to PIR:H64316 percent identity: 44.847 in
382aa"
/codon_start=1
/transl_table=11
/product="383aa long hypothetical protein"
/protein_id="BAA29398.1"
/db_xref="PID:d1030341"
/db_xref="PID:g3256715"
/db_xref="GI:3256715"
/transl_table="MIVMDREELIILEKYDKDKITIGVIGSHSMDIADAKKEGF
PVLVVAQRGRHRTYAKYFKLRKTKDNLTGFIIDEVITILKRCOITLQELIKRNVIF
INRSFVVTGIDKVENEFKVPMEGTRSLRTEEREKESYVWLEKALPYEEVKP
EEDIDGVALVWKLPAKKRLEGRFTAAAYKSEKSESLILGVITREDLKKARER
YIIGVFENFDEIISPIDEIEILGVMDRETSIDGIVRPAQOQLTPEHPEPEYTV
CGHASATLRESLEKVDMAEKYEVATKXVYPGIIIGPTLTQVVDKDLNFIIPVAP
RGGGNTIHMSGHVYNGNSLMSPTSGTRILNLEIKRGIEDELEKLYT"
/7392. .8486
/gene="PH0325"
/7392. .8486

```

Query Match 100.0%; Score 10; DB 1; Length 257000;
Best Local Similarity 100.0%; Pred No. 1.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
|||||
Db 235690 ACCCACACAG 235699

RESULT	5	ASRRNA16S/C	LOCUS	ASRRNA16S 1417 bp DNA	BCT	12-FEB-1997
DEFINITION	Actinomyces sp. 16S rRNA gene.					
ACCESSION	X81063					
NID	91838950					
VERSION	X81063.1 GI:1838950					
KEYWORDS	16S ribosomal RNA.					
SOURCE	Actinomyces sp.					
ORGANISM	Actinomyces sp.					
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.					
AUTHORS	1 (bases 1 to 1417)					
TITLE	Ramos,C.P., Foster,G. and Collins,M.D. Phylogenetic analysis of the genus Actinomyces based on 16S rRNA gene sequences: description of Arcanobacterium phocae sp. nov., Arcanobacterium bernardiae comb. nov., and Arcanobacterium pyogenes comb. nov.					
JOURNAL	Int. J. Syst. Bacteriol. 47 (1), 46-53 (1997)					
MEDLINE	97148975					
REFERENCE	2 (bases 1 to 1417)					
AUTHORS	Pascual,C.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-AUG-1994) C. Pascual, AFRC Institute of Food Research, Barley Gate, Welteknights Road, Reading RG6 2EF, UK					

FEATURES
Source
Location/Qualifiers
1..1417
/organism="Actinomyces sp."
/strain="ATCC 49338"
/db_xref="taxon:29317"
1..1417
/product="16S ribosomal RNA"
273 a 339 c 506 g 299 t
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 1417;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 440 ACCCACACAG 431
|||||
|||||

JUNIT 6
APPRIREP 4638 bp DNA BCT 26-JUL-1995
LOCUS A. rhizogenes plasmid pR14b replicator region (hairy-root-inducing
DEFINITION plasmid).
ACCESSION X04833
X04833
X04833.1 GI:39113
VERSION autonomous replication; origin of replication; plasmid; repA gene;
KEYWORDS repB gene; repC gene; R1 plasmid; unidentified reading frame.
SOURCE Agrobacterium rhizogenes.
ORGANISM Agrobacterium rhizogenes.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Agrobacterium.
1 (bases 1 to 4638)
REFERENCE Nishiyuchi, R., Takamami, M. and Oka, A.
TITLE Characterization and sequence determination of the replicator
JOURNAL Mol. Genet. 206, 1-8 (1987)
COMMENT Data kindly reviewed (23-SEP-1987) by Oka A.
FEATURES
Source
Location/Qualifiers
1..4638
/organism="Agrobacterium rhizogenes"
/plasmid="pR14b"
/strain="a transjugant of cured C58 and A. rhizogenes A4"
/db_xref="taxon:359"
1..164
/note="put. ORF1 (54 AA) (1 is 2nd base in codon)"
/codon_start=3
/transl_table=11
/protein_id="CAA28526.1"
/db_xref="PID:g809649"
/db_xref="GI:809649"
/db_xref="SPTREMBL:Q44434"
/translation="ILOGRIRLSRKGDFATIGRRRAGDLFLGYDCRITTKTTHETIQ
MAISLRHWK
complement(217..420)
/note="put. ORF 4 (AA 1-204)"
/codon_start=1
/transl_table=11
/protein_id="CAA28528.1"
/db_xref="PID:g80731"
/db_xref="GI:580731"
/db_xref="SPTREMBL:Q44435"
/translation="MLEVSGRGRHDAFSPRSTLQALLLGQTSPTQTHSLVAIF
AKLEKPRMOLAILLEGGELLG"
308..313
/note="put. rRNA binding site"
324..1538
/note="put. repA (AA 1-404)"
/codon_start=1
/transl_table=11
/protein_id="CAA28527.1"

/db_xref="PID:g39114"
/db_xref="GI:39114"
/db_xref="SWISS-PROT:P05682"
/translation="MAKSVLKAPVAVVGTALMERHADALSSLOAHHLKVPPEHSEK
GIRTFGPSEASKLIGVESYLRQTAEMPELNTVMSGPGGRFMSIDIHVIRKYMQV
GRGNRYLTPRRRGEOLQVYVNMFGSGGKTTAAHQAQYLAHGYRVYALDLDPOA
SLSALGSOPEMDVGNENLYGAIRYDDQVATIERVVRGTYPDLHLIGNLEMEFE
HDPRRLMNRKEDTFLFYGRISQYIEDIADNDVYVIDCPQIGYITLSALTAITSIL
VTVHPQMLDVMSNQLFAMTSNLRLEINAGAKFKFMKRYLITREPSPDQPNQVNG
YLRISGENYLNPFMLKTTAVSDAGLTNQLFEVERGLFTSTYDRLALEMNAVNDEI
EILIRKAWGRPT"
complement(426..429)
/note="put. rRNA binding site"
complement(585..941)
/note="put. ORF 5 (AA 1-119)"
/codon_start=1
/transl_table=11
/protein_id="CAA28529.1"
/db_xref="PID:g580732"
/db_xref="GI:580732"
/db_xref="SPTREMBL:Q44436"
/translation="MEGVNVSDDSDFCCHLIIIPYSAVERFVRANVRLAPKECR
KARLRIEIGODISAHSEVLRQMRGGGLTRPTFEIHRDNKLKLAAGTMQVAVP
ADLRHILPNHNDIFN"
complement(947..950)
/note="put. rRNA binding site"
complement(1510..1620)
/note="put. ORF 6 (AA 1-36)"
/codon_start=1
/transl_table=11
/protein_id="CAA28531.1"
/db_xref="PID:g39118"
/db_xref="GI:39118"
/db_xref="SPTREMBL:Q44437"
/translation="MERFPAVLLSAGDVSGASVETPRMCFRLMNAVPLF"
1606..1610
/note="put. rRNA binding site"
1619..2537
/note="put. rep B (AA 1-312)"
/codon_start=1
/transl_table=11
/protein_id="CAA28530.1"
/db_xref="PID:e189998"
/db_xref="PID:g1333706"
/db_xref="GI:1333706"
/db_xref="SWISS-PROT:P05683"
/translation="MPLGVTRKRRDPATKLTANTGNALRQNDRLSPAEIERELAE
GQAVITLDASSIETSPYODRMGDDIDGLTSTRBQGOVPIIVRPHPSODGRYOYAFG
HRLRAVSELGLPRAVARELTDEQVYVAQGGENREDLTEIKARFNHRLNRQFSR
EIVIAMSIDKSNLSMLLVLPSELDAIGAAGVGRPMOOLAEELIEVYSSPAD
VAKYAMSEVQALPSAERKAVIASLSPRVARGPEVATPDGRIAQVOTSKAKLE
ITDRKATPDFAFVLDHVPALYQAVHANGKRKE"
complement(1625..1628)
/note="put. rRNA binding site"
2670..2705
/note="put. rRNA binding site"
2710..3927
/note="put. rep C (AA 1-403)"
/codon_start=1
/transl_table=11
/protein_id="CAA28532.1"
/db_xref="PID:g39119"
/db_xref="GI:39119"
/db_xref="SWISS-PROT:P05684"
/translation="MQGTGSVTPPGRPRMTALVPRQIALADITOGKTAQKAKYFRD
ASAPMELIGTQNSLAVIDALSTFHPERRELQEAQIVFSSNAQALARHAGAGATLR
RHIAMLVESGLIVRKDQSAKGRYARKDQGIERAFGDLSPILARSELAMAOQVM
ADRAAFPMAKESLITCRDVRKLTITAMEGEGQMOAVEYVELVGRIPAPPLAD
VSILSEEMNLOEIIINRLIENNSNTNAOSHOIIONSPEPSVNELEPRSEKQ
GAKPSIDARAREPTIKAPRLGWTIKACPTIGVSGSAVSRDIMSAAVYVRSILGQ
SPSNVODACENAGPEMAAAMACILERANFINSPGGLYLDLTRSELGFFSLGPMIMA
LKASGQGLTRFG"
3951..3960
repeat_unit


```

repeat_unit /note="imp. inverted repeat A"
              3961..3976
              /note="imp. inverted repeat A'"
RBS          3980..3986
              /note="put. rRNA binding site"
CDS          3992..4255
              /note="put. ORF 2 (AA 1-87)"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA28533.1"
              /db_xref="PID:399120"
              /db_xref="GI:39120"
              /db_xref="SPTREMBL:Q44438"
              /translation="MSVFMIDRPSDSSTRKALGAILTLNKHGCPMPYIDRESEPA
              LHAIVFCRRHVAHCLEFALGIVDSITELADKAITHTEDD"
              4255..>4638
              /note="put. ORF 3 (AA 1-128)"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA28534.1"
              /db_xref="PID:399121"
              /db_xref="GI:39121"
              /db_xref="SPTREMBL:Q44439"
              /translation="MADEIOWPDGTAAAEVPAAPGRRAKAPAPETARQPSKVS
              KPTKGLSNREKLEKIGQIEAOVAGATLKDAVKIVGISVQIYOMKRAAVQPVSONP
              AVSVSDDELGFIOLEENRKLRLK"
BASE COUNT  1129 a 1191 c 1348 g 970 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 4638;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 4235 ACCCACACAG 4244

RESULT 7
LOCUS BACCMCASE 1928 bp DNA BCT 02-FEB-1999
DEFINITION B. subtilis carboxymethyl cellulase (CMCase) gene.
ACCESSION D01057
NID 9216387
VERSION D01057.1 GI:216387
WORDS CMCase.
SOURCE Bacillus subtilis (strain:BSE616) DNA.
ORGANISM Bacillus subtilis
          Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 1928)
AUTHORS Park,S.H., Kim,H.K. and Pack,M.Y.
TITLE Characterization and structure of the cellulase gene of Bacillus
          subtilis BSE616
JOURNAL Agric. Biol. Chem. 55 (2), 441-448 (1991)
MEDLINE 9129280
COMMENT Submitted (10-JUN-1991) to DDBJ by:
          Seung H. Park
          Genetic Engineering Research Institute
          P.O. Box 17
          Taejeok Science Town
          Taejeon, 305-606
          Korea
          Phone: 042-829-4312.
          Location/Qualifiers
            1..1928
            /organism="Bacillus subtilis"
            /strain="BSE616"
            /db_xref="taxon:1423"
            -35..signal
            -10..signal
            RBS 209..215

```

```

CDS          228..1727
              /note="CMCase"
              /codon_start=1
              /transl_table=11
              /protein_id="BAA00859.1"
              /db_xref="PID:d1001323"
              /db_xref="PID:9216388"
              /db_xref="GI:216388"
              /translation="MKRSISLFTCLLTLLTMGMLASPAAGTKTPVAKNQSLT
              KGTOLVNRGKAVOLKGISHLQWYGEYVKRDSIKMLRDMGKITVFRAAMYADGGI
              IDNPSVKKMKAEVAEAKELGIYIIDHILINDGNPNONREKAEFEKMSLTYGNT
              NVIYIANEPNGDVNMKRDIKPYAEVLSVRKNDPNIITVGTGMSQVNDADDO
              LKDAVMDALHFAVAGTGOEFLRDRANVALSGCAPLVPMGCTSPASGSGVFLQORE
              WKIKLDSFTISVWMNNSDKROESSALKPGSKTGGFRFLSLSASGTFVRNITCTKD
              STKIDPETPADKPTQENGISVQYRAGDGSNNSNOIRPQLOIRKNGNTTVDLKVTA
              RYVNAKNKGNQVNDVYOLGCGNVTYKRVTLHKRPGQADTYLELGFKNKGLTAPGASTG
              NIQLRHNDQWSNVAQSGDYFFKSNFFKTKITLYDQKLIWGTBN"
              1743..1750
              /rpt_type=inverted
              /rpt_type=1764
              /rpt_type=inverted
BASE COUNT  661 a 360 c 452 g 455 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 1928;
Best Local Similarity 100.0%; Pred. No. 3.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
Db 1268 ACCCACACAG 1277

RESULT 8
LOCUS BC170DEGR 26170 bp DNA BCT 24-JUN-1998
DEFINITION B. subtilis DNA (26.2 kb fragment; 170 degree region).
ACCESSION Z73234
NID 91405443
VERSION Z73234.1 GI:1405443
WORDS acnltase; alst gene; ccdA gene; endo-1,4-beta glucanase;
          endo-1,4-beta xylanase; grlA gene; grlB gene; tlpA gene;
          transketolase; yneA gene; yneB gene; yneC gene; yneI
          gene; yneJ gene; yneK gene; yneL gene; yneP gene; yneQ gene; yneR
          gene; yneS gene; yneT gene; ynfC gene; ynfE gene; ynfF gene.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
          Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 26170)
AUTHORS Rose,M. and Entlan,K.D.
TITLE New genes in the 170 degrees region of the Bacillus subtilis genome
          encode DNA gyrase subunits, a thiorodoxin, a xylanase and an amino
          acid transporter
JOURNAL Microbiology 142 (Pt 11), 3097-3101 (1996)
MEDLINE 97124194
COMMENT Direct Submission
          Submitted (19-JUN-1996) Rose M., Johann Wolfgang
          Goethe-Universitaet Frankfurt, Institut fuer Mikrobiologie,
          Marie-Curie-Str. 9, Geb. N250 Frankfurt/M. GERMANY D-60439
          Overlapping sequence: X87845 (bases 3016-6779).
          Location/Qualifiers
            1..26170
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
            82..399
            /gene="yneA"
            82..399
            /gene="yneA"

```



```

/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97614.1"
/db_xref="PID:e249640"
/db_xref="PID:g1405444"
/db_xref="GI:1405444"
/db_xref="GI:1405444"
/db_xref="SPTREMBL:045056"
/translation="MIMSKESTIIFGLFTVILASVILMISTSSQCELNQYKIEVQO
GDTLMSIADQVADTKRKINKNDIEIWWADKNQLOTSDIQPGDELVIPLKKHODAYELA
TVR"
418..1071
/gene="ynef"
418..1071
/gene="ynef"
/feature="similar to gamma-delta resolvase"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97615.1"
/db_xref="PID:e249641"
/db_xref="PID:g1405445"
/db_xref="GI:1405445"
/db_xref="SPTREMBL:045057"
/translation="MKALITAVSTNKQOENSLKROEBELTAIENGMEVYKISE
KASGIEMDRGVELDELKKNADIVILVQDETIRGRGNKAKTALHLCYREGVYVTT
AHRGELSEADSVLEIVSYVEEQRIHNNKIRGKRAVKNQFPOKNLKNQCHN
SGKEIEVPISSIVLRANKLTFEAIATLRNGYDVSATVHRFOYIENEETAE"
1537..3540
/gene="ckta"
1537..3540
/gene="ckta"
/codon_start=1
/transl_table=11
/product="transketolase"
/protein_id="CAA97616.1"
/db_xref="PID:e249642"
/db_xref="PID:g1405446"
/db_xref="GI:1405446"
/db_xref="SWISS-PROT:P45694"
/translation="MDTIEKRSVATIRLSTDAIRKANGSHGPMGAAPNAVITLTK
FMVSPANGMGNRDRFVLSAGHSALYSMTLHSGPDLSTEDLKGFQWQSKTGHF
EGHRNAGDATGTPGGGIMAVGKALIERHIAETYNDSFVNDVHTYISCGDDLM
EGISSEASLAGHLQGRILVYDSNDLSDDLDLRSSENVKQFEEMNEVLVED
GNNIELTAIEKARONKPKPLIEVKTITIGFSPNAGTSGVHGAPLAESESKLKE
AYATVEEDFVPSSEVYEHFAVAKESEKKEQWNAQFAKKEVYPELAOLEIAIK
GELPKDMQOEAVYVEKSSLSRASSGEVLNGLAKIPFVGGASDLAGSNKTTIKNA
GDTFVDSYSGKNMFCVRREPANGALNGLHGLRVRFGCTFFYSDLRPAIRLAL
MGLPTVYFTHDSIIVGEDGPTHEPVEQLASIRAPNLSLRPADGNETAAMKLAQV
STDPTALVLTFRONLPTIDQISEALAVKGAAYVSKSKNETPDALLIAGSSEVGLA
IEQAELAKENIDVSVMPSPMDRERKQSDKNEVLPAVYKRLAIEMGSSFGWGY
TGLEGDVIGDIRFGASAPETIINEYGFSPVVRVAKALINK"
3693..4139
/gene="ynef"
3693..4139
/gene="ynef"
/feature="identical to yoxf (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97617.1"
/db_xref="PID:e249643"
/db_xref="PID:g1405447"
/db_xref="GI:1405447"
/db_xref="SWISS-PROT:P45707"
/translation="MERNYTYTLIKEEPANHYFGESVYFELFQDYHTSLEKQYEM
IEKQIQTITPILHMHORLNMNKNKTDYQDLDITRIALPKAGHATFMKKEHME
IVASGDYEAETIFFEVLKRVSPCLAMDPSKRYGMLNPKVERNEV"
4225..4443
/gene="ynef"
4225..4443
/gene="ynef"
/feature="identical to yoxg (from acc. no. X87845)"

```

```

/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97618.1"
/db_xref="PID:e249644"
/db_xref="PID:g1405448"
/db_xref="GI:1405448"
/db_xref="GI:1405448"
/db_xref="SWISS-PROT:P45708"
/translation="MTLWVGILVGVALLIGVALGFFIARKYMSYLKNPINEOML
RMMAMQMKRPSQKINQMKAMNNQTK"
4910..5617
/gene="ccda"
4910..5617
/gene="ccda"
/feature="cytochrome C-type biogenesis protein"
/codon_start=1
/transl_table=11
/product="ccda"
/protein_id="CAA97594.1"
/db_xref="PID:e249645"
/db_xref="PID:g1405449"
/db_xref="GI:1405449"
/db_xref="SWISS-PROT:P45706"
/translation="MGDVNFTITGAGLSFISPCCLPLVPATLSYTGVSMDVKE
KLILQKRSLEFTLGLFVSIFIALGTSISFISLRDHDALROIGALLIIFGTI
TLGVFRPAMKERRIHERKHPSCGLSVLIGMAFPAAGWPCPTPIIAAVITLGTNP
GSAPVPMALYVLGAVPEFLISFTIKLKWIRKNQFLMKAGVLMIVIGVLFPMNM
SLIILSLDIEGGTGT"
5706..6068
/gene="ynef"
5706..6068
/gene="ynef"
/feature="response regulator"
/feature="identical to yoxh (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97595.1"
/db_xref="PID:e249646"
/db_xref="PID:g1405450"
/db_xref="GI:1405450"
/db_xref="SWISS-PROT:P45709"
/translation="MTRVLVVDKFRKVRKIRELLEBANTIIAGEADGQADLYRK
LRPLVMDITMPVKNIGIKALRDILTFDPRAKVIIMCTAMQOQRIVTEBAIEGADFIY
KPFETVLEAVSRVMGH"
6147..6638
/gene="ynef"
6147..6638
/gene="ynef"
/feature="identical to yoxi (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97596.1"
/db_xref="PID:e249647"
/db_xref="PID:g1405451"
/db_xref="GI:1405451"
/db_xref="SWISS-PROT:P45710"
/translation="MISMTIIVSIIIVLMAVAVVTRIKSSDKPSPKIIILPPIIM
STGLAMFLFVFWTGAELFAELFGLVIFSLIKTSFEIKNNEIYMKRSKAFVTL
VGLVIRIVKMSIISTSIDGALSGMFIAGFVIRIAMIYLSYRKLNELOSSNI
QMN"
6669..7097
/gene="ynef"
6669..7097
/gene="ynef"
/feature="identical to yoxj (from acc. no. X87845)"
/codon_start=1
/transl_table=11
/product="ynef"
/protein_id="CAA97597.1"
/db_xref="PID:e249648"
/db_xref="PID:g1405452"

```



```

gene
CDS
    /db_xref="GI:1405452"
    /db_xref="SWISS-PROT:P45711"
    /translation="MLEGFWLFIEMVIMVVLISIGFEWERELKRLPREDGRSE
    LPMODYIENSRLNDEKOFLELTAPEPELEFPAAKAKIAGKIGELALKENYAKI
    DOOLMKYIILAPKRDHTEFLKRLDKRIDLEPTOLIK"
    complement(7331..7723)
    /gene="yneL"
    /complement(7331..7723)
    /note="similar to heat shock proteins"
    /gene="yneL"
    /note="yneL"
    /codon_start=1
    /transl_table=11
    /product="yneL"
    /protein_id="CAA97598.1"
    /db_xref="PID:6249649"
    /db_xref="PID:91405453"
    /db_xref="GI:1405453"

Query Match      100.0%; Score 10; DB 1; Length 26170;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 23341 ACCCACACAG 23350

RESULT 9
LOCUS BSL1GLUC 2314 bp DNA BCT 30-JAN-1995
DEFINITION B. subtilis gene for cellulase.
ACCESSION X67044
VERSION 939776
KEYWORDS cellulase; endo-1,4-beta-glucanase.
SOURCE Bacillus subtilis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 2314)
AUTHORS Lindahl,V., Aa,K. and Tromsø,M.A.
TITLE Direct Submision
JOURNAL Submitted (22-JUN-1992) V. Lindahl, Agricultural University of
MEDLINE Norway, Dept of Biotechnological Sciences, PO Box 40, 1432 AS,
FEATURES
    source
        1 (bases 1 to 2314)
        /organism="Bacillus subtilis"
        /strain="CK-2"
        /db_xref="taxon:1423"
        /clone_lib="pGEM CEL 101"
        719..724
        742..747
        789..794
        807..2306
        /EC_number="3.2.1.4"
        /note="cellulase"
        /codon_start=1
        /transl_table=11
        /product="endo-beta-1,4-glucanase"
        /protein_id="CAA47429.1"
        /db_xref="GI:39777"
        /db_xref="SWISS-PROT:P10475"
        /translation="MKRSISFICLLITLTMGSMISPSAAGTKTPVAKNQLSI
        KGTOLYNDRKAVOLCKGISSHGLQYTGTVKDSIKRLDMGILTFRAAYTADGGT
        IDNFSVKNKVAEAAKELGIYIIDWHILINDGNPNONKEKAEFFKEMSLYGNTP

```

```

Query Match      100.0%; Score 10; DB 1; Length 2314;
Best Local Similarity 100.0%; Pred. No. 3.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
Db 1847 ACCCACACAG 1856

RESULT 10
LOCUS BSBGLUC2 2435 bp DNA BCT 30-MAR-1995
DEFINITION Bacillus subtilis endo-beta-1,4-glucanase gene.
ACCESSION X04689
VERSION 939823
KEYWORDS beta-glucanase; endo-beta-1,4-glucanase; glucanase; unidentified
          reading frame.
SOURCE Bacillus subtilis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
          Bacillus.
REFERENCE 1 (bases 1 to 2435)
AUTHORS Mackay,R.M., Lo,A., Willick,G., Zuker,M., Baird,S., Dove,M.,
TITLE Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene
JOURNAL Nucleic Acids Res. 14 (22), 9159-9170 (1986)
MEDLINE 87066783
FEATURES
    source
        1..2435
        Location/Qualifiers
        /organism="Bacillus subtilis"
        /strain="PAP115"
        /db_xref="taxon:1423"
        <1..428
        /note="unidentified reading frame (142AA) (1 is 2nd base
        in codon)"
        /codon_start=3
        /transl_table=11
        /protein_id="CAA28391.1"
        /db_xref="PID:9757741"
        /db_xref="GI:757741"
        /db_xref="SPTREMBL:Q45525"
        /translation="IOVTOALNHHHIGNAPFIFVAMFLFAFSSVGNYYGETNIE
        FIKTSKTLNIVYIAVIAVAVVIGSLSGFOIWMADLFEMALINLIVIALSNVAY
        KYVKDYAKORKOGIDPYFKAKNIPGLKNAIWEDEKGA"
        443..473
        /note="put. transcription terminator"
        443..456
        /note="inverted repeat A"
        460..473
        /note="inverted repeat A"
        643..648
        /note="put. -35 region"
        666..671
        /note="put. -10 region"
        712..726
        /note="put. rRNA binding site"
        731..817
        /product="signal peptide (AA -29 to -1)"
        731..2230
        /note="precursor polypeptide (AA -29 to 470)"
        /codon_start=1
        /transl_table=11

```



```

/protein_id="CAA28392.1"
/db_xref="PID:g39824"
/db_xref="GI:39824"
/db_xref="SWISS-Prot:P10475"
/translation="MKRSISIFITCLITLTLTLMGMLASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALKGISSHGLQWYGEYVKNDSIKMLRDMGTFVPAAYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHLINDGNPNONKEKKEFEKMSLYGNTP
NVIETINERNGDVNMRKDIKPAVEEIVSIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISWVNMNLSDKQESSALKPGASKGWRLSDSASGTFVENILGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSNOIRPOLIKNNGNITVDLKQVTA
YWKAKKNGQNFCDYAOIGGNTVHKFVTLHKPKQAGDTLELGFKNKGTLPAGASTG
NIOLRLHNDMSVNTAAGSDYSFKNSTFKTKITLIDGKLLMGTEPN"
mat_peptide
/product="endo-beta-1,4-glucanase (AA 1-470)"
repeat_unit
/translation="2247..2254"
/feature="inverted repeat B"
terminator
/feature="put transcription terminator"
repeat_unit
/feature="inverted repeat B"
BASE COUNT      819 a      450 c      556 g      610 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 2435;
Best Local Similarity 100.0%; Pred. No. 3.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 ACCGACAG 10
Db      1771 ACCGACAG 1780
RESULT 11
BSEQLSG      1800 bp      DNA      BCT      26-MAR-1998
LOCUS      B.subtilis (168 trpC2) egIs gene for endo-1,4-beta-glucanase.
ACCESSION      229076
VERSION      9509266
KEYWORDS      egIs gene; endo-1,4-beta-glucanase.
SOURCE      Bacillus subtilis.
ORGANISM      Bacillus; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
REFERENCE      1 (bases 1 to 1800)
AUTHORS      Mackay,R.M., Lo,A., Willick,G., Zuker,M., Baird,S., Dove,M.,
TITLE      Moraneli,F. and Seligy,V.
JOURNAL      Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene
MEDLINE      Nucleic Acids Res. 14 (22), 9159-9170 (1986)
REFERENCE      2 (bases 1 to 1800)
AUTHORS      Borriass,R.
TITLE      Direct Submission
JOURNAL      Submitted (14-DEC-1993) Borriass R., Humboldt Univ. zu Berlin,
MEDLINE      Biologie, Warschauerstrasse 43, Berlin, Deutschland, D-10243
REFERENCE      3 (bases 1 to 1800)
AUTHORS      Wolf,M., Geczi,A., Simon,O. and Borriass,R.
TITLE      Genes encoding xylan and beta-glucan hydrolasing enzymes in
JOURNAL      Bacillus subtilis: characterization, mapping and construction of
MEDLINE      Microbiology 141 (Pt 2), 281-290 (1995)
FEATURES
SOURCE      Location/Qualifiers
              1..1800
              /organism="Bacillus subtilis"
              /strain="168 trpC2"
              /db_xref="taxon:1423"
              280..1779
              /gene="egIs"
              280..1779
              /gene="egIs"
              /gene="egIs"
CDS

```

```

/standard_name="1,4-beta-D-glucan-4-glucanohydrolase"
/ec_number="3.2.1.4"
/function="hydrolysing carboxymethyl cellulose"
/note="Sequence has been found completely identical with
that published by R.M.McKay et al. (1986) NAR 14,
9159-9170 from B.subtilis Pap115 (nt452-nt2251)"
/codon_start=1
/label="table=11"
/evidence="experimental"
/product="endo-1,4-beta-glucanase"
/protein_id="CAA2317.1"
/db_xref="PID:g509267"
/db_xref="GI:509267"
/db_xref="SWISS-Prot:P10475"
/translation="MKRSISIFITCLITLTLTLMGMLASPAAGTKTPVAKNGOLSI
KGTOLVNRDGAVALKGISSHGLQWYGEYVKNDSIKMLRDMGTFVPAAYADGGY
IDNPSVKNKVEAVEAKELGIYIIDMHLINDGNPNONKEKKEFEKMSLYGNTP
NVIETINERNGDVNMRKDIKPAVEEIVSIRKNDPNIIIVGTGTSQDVNDADDO
LKDNVMTALHFAAGTHGQFLRDKANALSKGAPFVTEWGTSDASGNGVFLDQSR
WLKYLDSKTSISWVNMNLSDKQESSALKPGASKGWRLSDSASGTFVENILGTRD
STKDIPTSPKDKPTQENGISVQYRAGDSMNSNOIRPOLIKNNGNITVDLKQVTA
YWKAKKNGQNFCDYAOIGGNTVHKFVTLHKPKQAGDTLELGFKNKGTLPAGASTG
NIOLRLHNDMSVNTAAGSDYSFKNSTFKTKITLIDGKLLMGTEPN"
BASE COUNT      609 a      336 c      416 g      439 t
ORIGIN
Query Match      100.0%; Score 10; DB 1; Length 1800;
Best Local Similarity 100.0%; Pred. No. 3.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 ACCGACAG 10
Db      1320 ACCGACAG 1329
RESULT 12
BSUB0010      233780 bp      DNA      BCT      26-NOV-1997
LOCUS      Bacillus subtilis complete genome (section 10 of 21): from 1781201
ACCESSION      299113 AL009126
VERSION      92634090
KEYWORDS      299113.1 GI:2634090
SOURCE      Bacillus subtilis.
ORGANISM      Bacillus; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
REFERENCE      1 (bases 1 to 233780)
AUTHORS      Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
TITLE      Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
JOURNAL      Borriass,R., Bounisier,L., Brans,A., Braun,M., Brignell,S.C.,
MEDLINE      Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
REFERENCE      Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
AUTHORS      Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
TITLE      Emerson,P.T., Entlan,K.D., Errington,J., Fabeet,C., Ferrara,A.,
JOURNAL      Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
MEDLINE      Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
REFERENCE      Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,
AUTHORS      Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
TITLE      Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
JOURNAL      Klaere-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
MEDLINE      Konigstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
REFERENCE      Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
AUTHORS      Masuda,S., Maue,C., Medigue,C., Medina,N., Melado,D., O'Reilly,M.,
TITLE      Mizuno,M., Moestl,D., Nakai,S., Noack,M., Noone,D., O'Reilly,M.,
JOURNAL      Ogawa,K., Ogihara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
MEDLINE      Portetelle,D., Porwollik,S., Prescott,A.M., Prescott,E., Pujic,P.,
REFERENCE      Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
AUTHORS      Rivolta,C., Rochat,B., Roche,B., Rose,M., Sadate,Y., Sato,T.,
TITLE      Scanlan,E., Schleich,S., Schroeter,R., Scottone,F., Sekiguchi,J.,

```


[illegible]

8255. .9292
/gene="pkgs"
/function="Involved in polypeptide synthesis"
/codon_start=1
/transl_table=1
/protein_id="CAB13597.1"
/db_xref="pid:el183372"
/db_xref="pid:g2634097"
/db_xref="pid:g2634097"
/db_xref="SWISS-PROT:P40830"
/transl_table="MLTTCSESGIDFGKSLTYIHEYILNRCNLEEVKQACSGTAA
GFOAVALFISOTSPGAKALVIASDISEFLLEAGDALSSEMSYAEPAGAAVAALV
GENEVEFDIDPGANGYGYEYMDCKRHPIDSEADSDSLTDCGOTLEFQKRRV
PGANVQDFPYGLAYHTPGVGNKAKHPRMKRVAKYKTSGETFPLFRKGLNCCQR
VGNMGAALFLALASTIDQGRFDTPKRIQCSYSSGCCSEYSGITTPQGEQRRTFG
IEKHLDRYQLSMEYEYELFRSGVNRGTNRVXLDLPMIPGIMOSTOEKRPFLLEET
SEFRKRWYS"
9280. .10059
/gene="pkgs"
9280. .10059

Query Match 100.0%; Score 10; DB 1; Length 233780;
Best Local Similarity 100.0%; Pred. No. 1,2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACCCACACAG 10
|||||
DB 159720 ACCCACACAG 159729

RESULT 13
BSUB0017
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BSUB0017 217420 bp DNA BCT 26-NOV-1997
BSUB0017 subfills complete genome (section 17 of 21): from 3197001
to 3414420.
299120 AL009126
92635613
299120.1 GI:2635613

Bacillus subtilis.
Bacillus subfills
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
1 (bases 1 to 217420)
Kunst, F., Oggensawa, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Batterso, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Bignon, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codan, J.J., Conneton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devigne, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emerson, P.T., Ertlan, R.D., Errington, J., Fabret, C., Ferrari, E.,
Folger, D., Fritz, C., Fujita, M., Fujita, Y., Funa, S., Galizzi, A.,
Gallardon, N., Ghm, S.Y., Glaser, P., Goffeau, A., Golligity, E.J.,
Grand, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer-Blanchard, M., Klein, C., Kobayashi, Y., Koertgen, P.,
Koenigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Ladinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Meuel, C., Medigue, C., Medina, N., Mellado, R.P.,
Minato, M., Mostl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
Punelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Riviere, C., Roche, E., Roche, B., Rose, M., Sadale, I., Sato, T.,
Schlan, E., Schleich, S., Schroeter, R., Scottone, F., Sekiguchi, J.,
Sewokan, A., Setor, S.J., Serior, P., Shin, B.S., Solido, B.,
Sotokan, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpieta, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenberg, M., Vannier, F., Vassart, A.,
Viert, A., Wandut, R., Wedler, E., Wedler, H., Weitzenger, T.,
Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasunaga, K., Yata, K.,
Yoshida, K., Yoshikawa, H., Yuzumstein, E., Yoshikawa, H. and

TITLE	gene	CDS
Danchin, A. The complete genome sequence of the gram-positive bacterium <i>Bacillus subtilis</i> Nature 390 (6657), 249-256 (1997) 98044033 2 (bases 1 to 217420) Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A. Direct Submission Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszerpasteur.fr, adanchinpasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48	Location/Qualifiers 1. .217420 /organism="Bacillus subtilis" /strain="168" /db_xref="taxon:1423" complement(3. .1277) /gene="yule" complement(3. .1277) /gene="yule" /function="unknown" /note="similar to L-rhamnose isomerase" /codon_start=1 /transl_table=1 /protein_id="CAB15107.1" /db_xref="PID:e1184196" /db_xref="PID:g2635614" /db_xref="GI:2635614" /db_xref="SPTREMBL:005264" /translation="MTIKANYDSAKQAYEKWGIIDVEBALRQLRQVPISIHCMQGDIE GVEENKELSGGIDVTGNYGKATQPELRRDLKALSLPGKRVNLAIVETNEE AVERDELKPOHFNWYMAKNIIGLDENPTLFSHEKAAGILSHSDPPIRPFWRH CIACRIRGEYFGKELGRCPLTNINIPGYDIOSDRLTPRKAKESIDRFSEISQ HNDISIKSLFGLGJSESIVVGSHEFTYALVLTNKKCLLDTGHPHPEYVSNISL LYNDKLALVSRVPMDSHVYVLDLRLAIEIYNNALAEVAIGLDFFDSINRV AAWTIGRNMIKALLVALLPGLIKQLEGEGRTERLALMEEFKTYFPGAIDSYCE OMAGPYEAMLYKIEYEQCVLLKRRASPIV" complement(1303. .1617) /gene="yuld" complement(1303. .1617) /gene="yuld" /function="unknown" /note="similar to hypothetical proteins" /codon_start=1 /transl_table=1 /protein_id="CAB15108.1" /db_xref="PID:e1184197" /db_xref="PID:g2635615" /db_xref="GI:2635615" /db_xref="SPTREMBL:005263" /translation="MKRKASIMFVHODKREYIQRHDDITPEMAEAKAGAHHSIE LDSETGLFAYLIEIDEKMKADTEVCCQRMWKSAPLMKNTSDFSVAIDKEVY LD"	complement(1635. .3092) /gene="yulc" complement(1635. .3092) /gene="yulc" /function="unknown" /note="similar to rhamnulokinase" /codon_start=1 /transl_table=1 /protein_id="CAB15109.1" /db_xref="PID:e1184198" /db_xref="PID:g2635616" /db_xref="GI:2635616" /db_xref="SPTREMBL:005262" /translation="MYTATADGASGRIMVGLNECKLIDIORIHPRANGSORGHC LMDIDLKQILOGLOKRVTLGSEHCYIGDTNAYDVYLLDERGDRIRREISDRRT DHTIDLEHLTKRAAIYOKTGIOFOPNTIYOLFEEDERLKRTDKIMAPIDVLGYCL TGAIVAEITVNSTQILLANSTGNLDEPLEAVSLDEQFAPLPEQCEGLKRNVEP DYDLACKEVKTATHTDASVIAAPGVNDMAVYISGTVSLIGVENKTPITPLAEN

NVTNRGANNTIRFLKNTIGMVIQVIOEVKQOLADYSEFOOLAEEAKKTEPFOOFLIND
 KRLINPENMIKEIOHYCROTROKIPRTAGELACCICYSLATIYAIKELTEITTEKPI
 EOHFIIGGARDFLNOITADSGKRVAGTIEATATENTLIMOTIAKEVADIKEARQ
 VVRNSFPKIVFTPKIDIRSTIIQSFQOTVALASK"
 complement(3097. .3873)
 /gene="yulB"
 /complement(3097. .3873)
 /gene="yulB"
 /function="unknown"
 /note="similar to transcriptional regulator (Deor family)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15110.1"
 /db_xref="PID:el184199"
 /db_xref="PID:g2635617"
 /db_xref="GI:2635617"
 /db_xref="SWISS-PROT:O05261"
 /translation="MLVAERQOKIYEIYNMSSIRVSELSDFVTEETIRDLKLE
 KEHLSRSHGAVSIOOKESEHSEFSEITVIEKKATAHAARVAGSDRIIDAST
 TAYWAKTLPDIETLVITNSKAATELSKNKNISVISGGLKSLSPADPLARSL
 ETIYVAKTFLSCGKFDVNGMSDSEMOALKKRIEISDOTILMAOSSKGNFEFSH
 IASLDVSRLLITDSGLDPASVKALEDKRVAVAVPLSKRG"
 complement(3930. .5999)
 /gene="yuxG"
 /complement(3930. .5999)
 /gene="yuxG"
 /function="unknown"
 /note="alternate gene name: yulA; similar to
 sorbitol-6-phosphate 2-dehydrogenase"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15111.1"
 /db_xref="PID:el184200"
 /db_xref="PID:g2635618"
 /db_xref="GI:2635618"
 /db_xref="SWISS-PROT:P40747"
 /translation="MKHIMDSERMAOLPKVEELVYRNSLIGSDRTYCMNGGNTSM
 KTEEDRERELVMMVWKGSSDLATMAKNAHPSGKLDDIPRLKROMPEBEVADIL
 SHCMIDSKHPPRSITLILHAFLEPHVDHPDAITISCCDNKQALAEIYGRFVW
 VPVYRGTLSKMAIEGVANNPHELVMEHGVATWGETSETCYOKTISIOAEY
 INDRINQHEVFGKRYQPLPEDKROILAGIMPVYRGVSEKMIISYDHDVLEF
 VNSVOAPALSOIGAACPDHLVTKRPVLYIDMNETODVHLADLISGVTEFTELO
 AVTFRNOODGOIFESAPRVLLIGIGVWNGKGYAMKSVGALYRRAIAMKATSL
 CQFVSLHNESEHYEYPLLEIKULIAPPEAFSRKVALITGGAGGIGSACRFALE
 GGVYADLVNIEGAOKIAGEINDAVGRAAVADVTKEDVQSAERAAALAGGID
 IGVNNAAGLATSPEDETSLEKMNLMNMTLGGYPLVAREAKRQKHONRGSMTFVS
 KNSYAGKNASAVSKALETHLARCIAEGEGEIRVNSVLPDAVLOGSAIMGSSWR
 EERAAVYGIPEPDLEHYRKRRTALLVNIYEPDIAEALFFASSRAEKTGCMITVDG
 VPAAFTR"
 complement(6109. .8125)
 /gene="tlpB"
 /complement(6109. .6124)
 /gene="tlpB"
 /complement(6137. .8125)
 /gene="tlpB"
 /codon_start=1
 /transl_table=11
 /product="methyl-accepting chemotaxis protein"
 /protein_id="CAB15112.1"
 /db_xref="PID:el184201"
 /db_xref="PID:g2635619"
 /db_xref="GI:2635619"
 /db_xref="SWISS-PROT:P39217"
 /translation="WGRFIQWKQPSISKPLIAFLAVLILPVGLVAFSYQAMNL
 DRELISAKGVNELSTLQNKLEBVKAIIDYEBVDKDLLEKNKTLLEKRYQT
 TLNDVGAIVYASDEBKRLKYPDGSPKPGPDGRWKAQVAERGOVFEPEPTDA
 TGDIVYISKOLKDSGVIALDNLDELVLASKRLKIGEGFAITTGKNTKIYHPIT
 KPGTSGDWTNOVSKRESEFEYFECKEKAFTNKLTGKTIAGIYPISELODS
 SPVANTYIILCVSIVIGIILITIRITPRKLVSTKAKISSGDIETLTHSKN
 EFGQLGEFNEASALRSVIGVLOTSENVASSSEBELTASAOQSKATEHTTLAIEQ
 QMGNEQSEKLETSSNLSQNEGISKVAQASSTIYSSIOSSSEAGSGEALVETG
 OKMTIDQVQKAEVAVKGLFETKSDITSILNVINGIADQNTMLNLAIEAARAGEG
 RGSVVAEEVRLAVQASDAKIEGLIOEIVREISTSLMFSQVNHVEKGLITDQ

TASFQIYEMTIOISGELONLNAVTEQOLSAGSOEVSSAVEDISAVAKESSAGIODIA
 ASAEQOLASHEEISSSAETIANNAAEELQDITTKFKIES"
 complement(8204. .8228)
 /gene="mcpA"
 /complement(8204. .10224)
 /gene="mcpA"
 /complement(8239. .10224)
 /gene="mcpA"
 /function="chemotaxis deficiency toward glucose and
 alpha-methyl-glucoside"
 /codon_start=1
 /transl_table=11
 /product="methyl-accepting chemotaxis protein"
 Query Match 100.0%; Score 10; DB 1; Length 217420;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ACCCACACAG 10
 Db 150765 ACCCACACAG 150774
 RESULT 14
 BTU04366
 LOCUS
 DEFINITION Bacillus thuringiensis japonensis Bulbul CryIII delta-endotoxin
 gene, complete cds.
 ACCESSION U04366
 NID 9532523
 VERSION U04366.1 GI:532523
 KEYWORDS
 SOURCE Bacillus thuringiensis
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/staphylococcus group; Bacillus.
 1 (bases 1 to 3483)
 Sato, R., Takeuchi, K., Ogilwara, K., Minami, M., Kaji, Y., Suzuki, N.,
 Horii, H., Asano, S., Ohba, M., and Iwahana, H.
 Cloning, heterologous expression, and localization of a novel
 crystal protein gene from Bacillus thuringiensis serovar japonensis
 strain bulbul toxic to scarabeid insects
 Curr. Microbiol. 28 (1), 15-19 (1994)
 94100786
 MEDLINE
 REFERENCE
 JOURNAL
 AUTHORS
 TITLE
 JOURNML
 COMMENT
 FEATURES
 source
 On Aug 27, 1994 this sequence version replaced gi:436836.
 1..3483
 Location/Qualifiers
 1..3483
 /organism="Bacillus thuringiensis"
 /strain="bulbul"
 /sub_species="japonensis"
 /db_xref="taxon:1428"
 /clone_lib="Library of K. Ogilwara"
 /note="serotype 23"
 1..3483
 /codon_start=1
 /transl_table=11
 /product="CryIII delta-endotoxin"
 /protein_id="Aa21119.1"
 /db_xref="PID:9532524"
 /db_xref="GI:532524"
 /translation="MSPNNQNEVEIIDAISPTSVSNISYRPIANDQNTLOMANYKD
 YLAKTESNAELSRNPGTFISADAVGTGIDYITISIGLIVLSEVSLSLGL
 LPSNNENWQITMNNRVEELIDKILDSVSRALADLANRIAVETYNQALNLEWRNP
 HSTRSALVYERGNAAEAILRTMNGSSQNNYETPLPYTAQASLHLVMDVOYIG
 KEMGYQNDLDLEKQGVSYTARSDHCQVWYAGLNLKGTGAKQWVDYRFRREN
 VMVLDVLAEPNDARYPLETNAELTRELFTDPGVSYVAGOSTLISWYDMIPALP
 SFTLENLAKPPFTLLDRIHNTYTSFRQNGTIEYINYNMGQALITISYIGSSFNKTS

GLVAGAEIDIPVCGNDIYRVVMTYIGRTNSLGVNVPVTFPSNNTOKTYSKRFQFAG
 GKITDSEGLTEYENQSYSHRVSYITSEIKSTGGVIGVAVIPFGWTHSASARNFI
 YATKISQIPINKASRTSGAVNWPQEBLIVGCVMKISGSGSVVILRVATDKAGSQ
 RYRIRIRIADBRAGKFTISSRSPENPATYASIAVINTNSTASLITYSTFAVESQPI
 NLISSGSRFDISITKEAGANLYIDRIEFIPVNTLFEAEEDLVAKRVNLFENE
 KDLQTSVTDYQVQANALIECLSDLEYLPEKRLMDVAEKARLVQARLIDDTGFN
 RINGNMGWGTSTGEVVEGDLFRDRLRLTSAREIDETETPTYLQOIDEISLKPYT
 RYKLGPISSODLEIKLIRHRANOIKVNPDLDPVRVNSCGGVDRCSQGYVDA
 NLALENGENGNSSDSHAFSHIDTGEIDLNTGTIYFKLPTTNGNATLGNLEFV
 EEPPLSGETLEMAQOQOQWODMARKRASEKTYAAKQIDRLPADYODOKLNSGV
 EMSDILLAQNLVQSIPIYVNDALPEIPGNMTSEFTETNLQOANLYDQNALIPGD
 FRNGLENNMNTSDVNTQOLSDTSVLVTPMNSQVSOQFTVQPYRYLVRYLTAKESVG
 DGVYIIRDGANOTETELTFENICDDDTGVLSTDTSYITKTYEFPSTEQWIDMSEREG
 VFNESEVLEVEEE"

BASE COUNT 1217 a 577 c 712 g 977 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 3483;
 Best Local Similarity 100.0%; Pred. No. 2.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAG 10
 |||||

Db 476 ACCCAGACAG 485

RESULT 15
 BUHARGSC 485 bp DNA BCT 26-JUL-1995
 LOCUS Buchnera aphidicola Arginyl tRNA synthetase (args) gene, 5' end of
 DEFINITION cds and 16S ribosomal RNA promoter region.
 ACCESSION 118930
 NID 9349610
 VERSION 118930.1 GI:349610
 KEYWORDS args gene; arginyl-tRNA synthetase.
 SOURCE Buchnera aphidicola DNA.
 ORGANISM Buchnera aphidicola
 Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 REFERENCE 1 (bases 1 to 485)
 AUTHORS Munson, M.A., Baumann, L. and Baumann, P.
 TITLE Buchnera aphidicola (a prokaryotic endosymbiont of aphids) contains
 a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene;
 sequence determination and promoter and terminator analysis
 JOURNAL Unpublished (1993)
 FEATURES
 Location/Qualifiers
 1..485
 /organism="Buchnera aphidicola"
 /db_xref="taxon:9"
 /note="host: Mindarus victoria"
 /complement(1..150)
 /gene="args"
 /complement(1..>150)
 /gene="args"
 /codon_start=1
 /transl_table=11
 /product="arginyl-tRNA synthetase"
 /protein_id="AA072381.1"
 /db_xref="PID:9349611"
 /db_xref="GI:349611"
 /translation="MNLVLEISIKKALIESAPLDCNPGIKSSPNQTSQYVNGI
 ISIAKK"
 334..339
 /note="deduced by comparison with E. coli rnb promoter
 region"
 355..360
 /note="deduced by comparison with E. coli rnb promoter
 region"
 382..390
 /standard_name="box A"
 /note="deduced by comparison with E. coli rnb promoter
 region"
 /function="box A is involved in antitermination of the rnb
 operon"
 misc-feature 406..413
 /standard_name="box C"
 /note="deduced by comparison with E. coli rnb promoter
 region"
 /function="box C might be involved in antitermination of
 the rnb operon"

BASE COUNT 204 a 39 c 48 g 194 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 485;
 Best Local Similarity 100.0%; Pred. No. 4.2e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAG 10
 |||||

Db 414 ACCCAGACAG 405

Search completed: September 17, 1999, 22:01:22
 Job time: 16265 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:33 ; Search time 425.19 Seconds
(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518a-89
Perfect score: 10
Sequence: 1 ACCCACACAG 10

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

abase : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	2181	1	003577	Fragment of sequen
2	10	100.0	1109	1	003852	Sequence encoding
3	10	100.0	2181	1	003668	Fragment encoding
4	10	100.0	384	1	N90572	Caprine repeat ele
5	10	100.0	448	1	N90573	Caprine repeat ele
6	10	100.0	2589	1	N90574	Caprine repeat ele
7	10	100.0	618	1	N90766	Sequence of the am
8	10	100.0	517	1	N91654	CDNA sequence of t
9	10	100.0	547	1	N91655	Synthetic tumour n
10	10	100.0	3086	1	N91815	Egr-1-contg. OC3.1
11	10	100.0	1310	1	N91469	Human lymphotoxin
12	10	100.0	453	1	N93191	Human lymphotoxin
13	10	100.0	925	1	N93196	Monkey recombinant
14	10	100.0	618	1	Q03289	New synthetic huma
15	10	100.0	3344	1	N81442	Clone encoding hum
16	10	100.0	10475	1	N80616	Genomic clone enco
17	10	100.0	618	1	N82251	Variant lymphotoxi
18	10	100.0	1077	1	N91628	Sequence of gamma-
19	10	100.0	1320	1	N80599	Encodes lymphotoxi
20	10	100.0	618	1	Q06155	CDNA encoding Lymph
21	10	100.0	2654	1	N70597	Sequence encoding
22	10	100.0	5522	1	N70062	Sau3A-EcoRI 5.4 kb
23	10	100.0	10475	1	N80643	Sequence of human
24	10	100.0	456	1	N70441	Sequence encoding
25	10	100.0	1310	1	N70439	Sequence encoding
26	10	100.0	492	1	N70440	Sequence encoding
27	10	100.0	447	1	N70111	Sequence encoding
28	10	100.0	951	1	N70255	Sequence of interf
29	10	100.0	1014	1	N70256	Sequence of interf
30	10	100.0	1966	1	N70316	Sequence encoding
31	10	100.0	1530	1	N70318	Sequence encoding
32	10	100.0	907	1	N70905	Human Kallikrein-1
33	10	100.0	738	1	N71049	Human Kallikrein-1
34	10	100.0	2160	1	N91624	Sequence encoding
35	10	100.0	682	1	Q13586	5' flanking sequen
36	10	100.0	5534	1	Q13574	Encodes amino acid
37	10	100.0	1863	1	Q14271	Ech gene, DNA enco
38	10	100.0	1863	1	Q14272	D-3 dopaminergic r
39	10	100.0	1337	1	N50046	Sequence encoding
40	10	100.0	15155	1	N50107	DNA encoding Facto
41	10	100.0	1109	1	Q20612	Murine macrophage
42	10	100.0	1109	1	Q20728	Murine macrophage
43	10	100.0	17041	1	Q21065	Genomic DNA of hum

ALIGNMENTS

c	44	10	100.0	626	1	Q22228	Lymphotoxin gene.
c	45	10	100.0	6614	1	X26905	6.6 kb transcript

RESULT	1
ID	003577
AC	003577 standard; DNA; 2181 BP.
DT	02-AUG-1990 (first entry)
DE	Fragment of sequence encoding Heymann nephritis antigen, gp330.
KW	Heymann nephritis antigen; HNA; gp330; familial
KW	hypercholesterolaemia; membranous glomerulonephritis; ds.
OS	Rattus sp.
PN	EP-358977-A.
PD	21-MAR-1990.
PF	19-AUG-1989; 115331.
PR	23-AUG-1988; US-235211.
PR	22-FEB-1989; US-313682.
PA	(GEHO-) Gen Hospital Corp.
PI	Smith JA, Raychowdhury R, Niles JL;
DR	WPI: 90-084851/12.
DR	P-PSDB: R05533.
PT	Cloned Heymann nephritis antigen, gp330, gene -
PT	used for diagnosis and in treatment of membranous
PT	glomerulonephritis, wounds and gastric ulcers.
PS	Disclousure; P; English.
CC	The gene can be used to transform a host cell providing functional
CC	derivatives, free from impurities.
CC	Sequence 2181 BP; 533 A; 541 C; 561 G; 526 T;

QY	1	ACCACACAG 10
Db	444	ACCACACAG 453

RESULT	2
ID	003852
AC	003852 standard; DNA; 1109 BP.
DT	23-AUG-1990 (first entry)
DE	Sequence encoding macrophage derived inflammatory cytokine, MIP-2.
KW	MIP-2; macrophage-derived inflammatory cytokine; heparin; ds.
OS	Mus musculus.
FH	Key
FT	cds
FT	47..346
FT	Location/Qualifiers
FT	mat-peptide
FT	125..346
FT	/*tag- a
FT	/*tag- b
PN	WO9002762-A.
PD	22-MAR-1990.
PF	1-SEP-1989; 003798.
PR	2-SEP-1988; US-240078.
PR	1-SEP-1989; US-399971.
PA	(CHIR-) Chiron Corp., (UYRO-) Rockefeller Univ.
PI	Wolpe SD, Cerami A, Sherry B, Olson-Tekamp PA;
DR	WPI: 90-115959/15.
DR	P-PSDB: R05790.
PT	Macrophage-derived inflammatory cytokine MIP-2 -
PT	used for developing screening assays, for prodn. of antibodies
PT	and in diagnosis and therapy.
PS	Claim 5; Fig 7; 78pp; English.
CC	MIP-2 may be used to detect idiopathic or invasive stimuli.
CC	Antibodies raised to the cytokine may be used to measure MIP-2
CC	activity and inhibit the production of MIP-2 used in treatment of
CC	inflammation or fever.

Sequence 1109 BP; 320 A; 228 C; 241 G; 320 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 1109;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
DB 735 ACCCACACAG 744

RESULT 3

ID 003668 standard; DNA; 2181 BP.
AC 003668;
DT 07-AUG-1990 (first entry)
DE Fragment encoding gp330, Heymann's nephritis antigen.
KW Heymann's nephritis antigen; gp330; LDL receptor;
Familial cholesterololemia; FXIIa; ss.

PN Rattus rattus.
WO9002207-A.
PD 08-MAR-1990.
PF 23-AUG-1989; U03621.
PR 23-AUG-1988; US-235211.
PR 22-FEB-1988; US-313682.
PR 22-AUG-1989; US-396697.
PA (GEHO-) General Hospital Corp.
PI Smith JA, Raychowdhury R, Miles JL;
DR WPI: 90-099425/13.
P-PSDB; R05580.

PT Recombinant DNA encoding and expressing Heymann's nephritis antigen -
its use for therapeutic and/or preventative administration.
PS Claim 3; Fig 8; 65pp; English.
CC The sequence was derived from two overlapping clones of 1.4 kb and a 2 kb
CC resp. It encodes three cysteine-rich regions which are homologous to the
CC LDL receptor binding domain.
Sequence 2181 BP; 533 A; 542 C; 580 G; 526 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2181;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
DB 444 ACCCACACAG 453

RESULT 4

ID N90572 standard; DNA; 384 BP.
AC N90572;
DT 12-MAR-1990 (first entry)
DE Caprine repeat element GRV.1a(a).
KW GRV.1a(a); Y-chromosome; ruminants; repeat element; sex determination;
OS Capra hircus.
PN WO8907154-A.
PD 10-AUG-1989.
PF 27-JAN-1989; AU00029.
PR 29-JAN-1988; AU-006476.

PA (ADRI-) Advanced Riverina Holdings Ltd.
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;
DR WPI: 89-249021/34.
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -
useful in hybridisation and polymerase chain reaction tests for sex
PT determination.
PS Fig. 8; 20 pp; English.
CC The sequence is isolated from male goats and hybridises only to the DNA
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal
CC DNA. It can be used for sex determination.

Sequence 384 BP; 113 A; 91 C; 87 G; 93 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 384;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
DB 155 ACCCACACAG 164

RESULT 5

ID N90573 standard; DNA; 448 BP.
AC N90573;
DT 12-MAR-1990 (first entry)
DE Caprine repeat element GRV.1b(a).
KW GRV.1b(a); Y-chromosome; ruminants; repeat element; sex determination;
OS Capra hircus.
PN WO8907154-A.
PD 10-AUG-1989.
PF 27-JAN-1989; AU00029.
PR 29-JAN-1988; AU-006476.

PA (ADRI-) Advanced Riverina Holdings Ltd.
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;
DR WPI: 89-249021/34.
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -
useful in hybridisation and polymerase chain reaction tests for sex
PT determination.
PS Fig. 9; 20 pp; English.

CC The sequence is isolated from male goats and hybridises only to the DNA
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal
CC DNA. It can be used for sex determination.
Sequence 448 BP; 132 A; 104 C; 103 G; 109 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 448;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAG 10
DB 159 ACCCACACAG 168

RESULT 6

ID N90574 standard; DNA; 2589 BP.
AC N90574;
DT 12-MAR-1990 (first entry)
DE Caprine repeat elements GRV.1a and GRV.1b.
KW GRV.1; Y-chromosome; ruminants; repeat element; sex determination;
OS Capra hircus.
PN WO8907154-A.
PD 10-AUG-1989.
PF 27-JAN-1989; AU00029.
PR 29-JAN-1988; AU-006476.

PA (ADRI-) Advanced Riverina Holdings Ltd.
PI Reed KC; Lord EA; Matthea KI; Mann DA; Beaton S; Herr CM; Matthews ME;
DR WPI: 89-249021/34.
PT New nucleic acid isolates hybridising only ruminant Y chromosomal DNA -
useful in hybridisation and polymerase chain reaction tests for sex
PT determination.
PS Fig. 10; 20 pp; English.
CC The sequence is isolated from male goats and hybridises only to the DNA
CC of cattle, sheep, goats and other ruminants which contain Y-chromosomal
CC DNA so can be used for sex determination.

Sequence 2589 BP; 754 A; 522 C; 548 G; 765 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2589;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
 DB 163 ACCCACACAG 172

RESULT 7

ID N90766 standard; cDNA; 618 BP.
 AC N90766;
 DT 13-MAY-1990 (first entry)
 DE Sequence of the amino acid-coding region of lymphotoxin cDNA
 KW Lymphotoxin; lymphotoxin-resistant cancer cell; antitumour
 FH activity.
 FT Key
 FT cds Location/Qualifiers
 FT 1..618
 FT /*tag= a

EP-314094-A.
 03-MAY-1989.
 26-OCT-1988; 117832.
 28-OCT-1987; JP-272034.
 (EISA) Eisai KK.
 PI Watabayashi T, Asada M, Nagasu T, Hasegawa Y, Shikata Y,
 PI Kuwada M;
 PI WPI; 89-131655/18.
 DR P-PSDB; P91909.
 PT Recombinant lymphotoxin derivs.
 PT - have improved antitumour activity or exhibit a
 PT cytotoxic effect on lymphotoxin-resistant cancer cells
 PS Figure 1; page 38; 57pp; English.
 CC A plasmid contg. cDNA coding lymphotoxin was obt'd. from a cDNA library
 CC from mRNA of Daudi cells.
 CC Sequence 618 BP; 110 A; 236 C; 132 G; 140 T;

Query Match 100.0%; Score 10; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. NO. 1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
 DB 547 ACCCACACAG 556

RESULT 8
 ID N91654 standard; cDNA; 517 BP.
 N91654;
 15-MAR-1990 (first entry)
 DE cDNA sequence of tumour necrosis factor beta.
 KW Tumour necrosis factor beta; cDNA.
 FH Key
 FT Location/Qualifiers
 FT cds 2..517
 FT /*tag= a

GB2217326-A.
 25-OCT-1989.
 PF 08-APR-1988; GB-008206.
 PR 08-APR-1988; 008206.
 PA (BBRI-) British Bio-Technology Ltd.
 PI Davies JA, Johnson ID;
 PI WPI; 89-311765/43.
 DR P-PSDB; P93080.
 PT Synthetic gene encoding tumour necrosis factor beta - incorporates
 PT restriction sites at frequent intervals to facilitate manipulation.
 PS Disclosure; Fig. 1; 18pp; English.
 CC The sequence is the deduced TNF beta cDNA sequence. It is prep'd.
 CC synthetically.
 CC Sequence 517 BP; 97 A; 199 C; 105 G; 116 T;

Query Match 100.0%; Score 10; DB 1; Length 517;
 Best Local Similarity 100.0%; Pred. NO. 1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
 DB 446 ACCCACACAG 455

RESULT 9

ID N91655 standard; DNA; 547 BP.
 AC N91655;
 DT 15-MAR-1990 (first entry)
 DE Synthetic tumour necrosis factor beta gene.
 KW Tumour necrosis factor beta.
 FH Key
 FT Location/Qualifiers
 FT cds 2..528
 FT /*tag= a
 FT 535..546
 FT /*tag= b

GB2217326-A.
 25-OCT-1989.
 PF 08-APR-1988; 008206.
 PR 08-APR-1988; GB-008206.
 PA (BBRI-) British Bio-Technology Ltd.
 PI Davies JA, Johnson ID;
 PI WPI; 89-311765/43.
 DR P-PSDB; P93081.
 PT Synthetic gene encoding tumour necrosis factor beta - incorporates
 PT restriction sites at frequent intervals to facilitate manipulation.
 PS Disclosure; Fig. 3; 18pp; English.
 CC The synthetic TNF beta gene has the following restriction sites: HindIII,
 CC BspHI, PvuII, NsiI, RsrII, KpnI, AccI, PflMI, PstI, BstXI, ScaI, BamHI
 CC and EcoRI. The claimed sequence is nucleotides 14-532. The TNF beta has
 CC antitumour activity. The gene can be easily modified due to the
 CC restriction sites. The flanking sites simplify incorporation of the gene
 CC into an expression system, eg mammalian cells, yeast and insect cell
 CC hosts. The gene is produced synthetically.
 CC Sequence 547 BP; 107 A; 206 C; 111 G; 123 T;

Query Match 100.0%; Score 10; DB 1; Length 547;
 Best Local Similarity 100.0%; Pred. NO. 1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
 DB 461 ACCCACACAG 470

RESULT 10

ID N91815 standard; DNA; 3086 BP.
 N91815;
 19-MAR-1990 (first entry)
 DE Egr-1 contg. OC3.1 clone.
 KW Egr-1; DNA-binding proteins; zinc fingers;
 KW early growth regulatory proteins.
 OS Mus.
 FH Key
 FT Location/Qualifiers
 FT mat.peptide 239..1857
 FT /*tag= a
 FT polya_signal 1865..1871
 FT /*tag= b
 FT polya_signal 3066..3072
 FT /*tag= c
 FT polya_signal 3055..3059
 FT /*tag= d

W08909777-A.
 19-OCT-1989.
 PF 07-APR-1989; U01473.
 PR 08-APR-1988; US-179587.
 PR 26-SEP-1988; US-249584.
 PA (ARCH-) Arch Development Corp.
 PI Sukhatne VP;
 PI WPI; 89-324195/44.

KW Synthetic human lymphotoxin (HL) gene.

Job time: 18973 sec

```

FH Key      Location/Qualifiers
FT cds      CDS                      1..618
FT          /*tag= a
FT          mat_peptide      103..618
FT          /*tag= b
PN J02000447-A.
PD 05-JAN-1990.
PF 27-OCT-1987; 27J513.
PR 27-OCT-1987; JP-27J513, JP-287035.
PA (SANY) Sanyo KK.
PI
DR WPI: 90-047988/07.
DR P-RSDB: R05232.
PT Human lymphotoxin synthetic gene
PT altered to leave only one restriction enzyme recognition site
PS Disclosure; Fig 1; 13pp; Japanese.
CC It is altered so that only one restriction enzyme site exists. It has
CC the same amino acid sequence but not the same base sequence as natural
CC HL and the addition of 30 nucleotides has eliminated the 6 bp palindromic
CC sequence. It makes modifications of HL easier and improved lymphotoxins
CC can be prep'd.
SQ Sequence 618 BP; 109 A; 238 C; 130 G; 141 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 618;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ACCCACACAG 10
DB 547 ACCCACACAG 556

```

```

RESULT 15
N81442/C
ID N81442 standard; DNA; 3344 BP.
AC N81442;
DE 17-OCT-1990 (first entry)
DE Clone encoding human protein S
KM Protein S; anti-thrombolytic agent; antibodies; ss.
OS synthetic.
FH Key      Location/Qualifiers
FT cds      247..2196
FT          /*tag= a
FT          /label-human Protein S
PN EP-255771-A.
   10-FEB-1988.
   24-JUL-1987; 306564.
   25-JUL-1986; US-890401.
PA (INTE-) Integrated Genetics.
PI Wydro R, Cohen E, Dackowski W, Stenflo J, Lundwall A, Dahlback B;
DR WPI: 88-037719/06.
DR P-RSDB: P81137.
PT Recombinant DNA encoding mature human Protein S - used therapeutically as
PT anti-thrombolytic agent and for producing antibodies for diagnostic use
PS Disclosure; P; English.
CC This clone encodes the complete mature human Protein S. To facilitate
CC expression in mammalian cells it is fused to the leader sequence of
CC bovine CDNA. See also N81441.
SQ Sequence 3344 BP; 1072 A; 575 C; 688 G; 1009 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 3344;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ACCCACACAG 10
DB 2925 ACCCACACAG 2916

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:00 ; Search time 2825.05 Seconds
(without alignments)

6.982 Million cell updates/sec

Title: US-08-956-518a-89
Perfect score: 10
Sequence: 1 ACCCACACAG 10

Scoring table: IDENTITY_NDC
Searched: 2546578 seqs, 98626752 residues

tabase :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	336	20	D20150	D20150 HUMGS01123
2	10	100.0	295	20	D22931	D22931 R1C1816A R
3	10	100.0	434	20	D24421	D24421 R1C18169A R
4	10	100.0	351	20	D24802	D24802 R1C18169A R
5	10	100.0	360	20	D36166	D36166 CELK029H4F
6	10	100.0	344	20	D39050	D39050 R1C18167A R
7	10	100.0	260	20	F01745	F01745 HSC0CB082 R
8	10	100.0	409	20	M78221	M78221 EST01814 Su
9	10	100.0	386	20	M78683	M78683 EST00831 H1
10	10	100.0	423	20	M79123	M79123 EST01271 Su
11	10	100.0	312	20	M79888	M79888 WEST00425 M
12	10	100.0	292	20	M85714	M85714 EST02233 Fe
13	10	100.0	323	20	M85791	M85791 EST02312 Fe
14	10	100.0	519	20	M88939	M88939 CEL1282 Chr
15	10	100.0	376	20	T03144	T03144 FB25F5 Feta
16	10	100.0	475	20	T03171	T03171 FB2C12 Feta
17	10	100.0	244	20	T03346	T03346 IB1300 InfA
18	10	100.0	407	20	T03378	T03378 IB1444 InfA
19	10	100.0	451	20	T03517	T03517 IB400 InfAn
20	10	100.0	376	20	T03683	T03683 IB742 InfAn
21	10	100.0	398	20	T03711	T03711 IB781 InfAn
22	10	100.0	442	20	T03812	T03812 IB934 InfAn
23	10	100.0	440	20	T03818	T03818 IB941 InfAn
24	10	100.0	250	20	T04879	T04879 EST02767 Fe
25	10	100.0	313	20	T07793	T07793 EST05683 Fe
26	10	100.0	237	20	T08188	T08188 EST06079 In
27	10	100.0	393	20	T09172	T09172 EST07065 In
28	10	100.0	212	20	T11879	T11879 A1093R Hear
29	10	100.0	400	20	T14435	T14435 SME0096 S
30	10	100.0	419	20	T14477	T14477 SME01038 S
31	10	100.0	442	20	T14830	T14830 CR8274 Lamb
32	10	100.0	454	20	T15430	T15430 IB1153 InfA
33	10	100.0	190	20	T15433	T15433 IB1181 InfA
34	10	100.0	318	20	T15490	T15490 IB1362 InfA
35	10	100.0	283	20	T15614	T15614 IB1633 InfA
36	10	100.0	271	20	T15962	T15962 IB2229 InfA
37	10	100.0	373	20	T16115	T16115 IB3508 InfA
38	10	100.0	180	20	T17312	T17312 NIB678 Norm
39	10	100.0	482	20	T23179	T23179 C2H2-604F H
40	10	100.0	319	20	T24025	T24025 seg2182 3HF
41	10	100.0	339	20	T29736	T29736 EST92350 Hu
42	10	100.0	389	20	T30661	T30661 EST20274 Hu
43	10	100.0	368	20	T31770	T31770 EST38563 Hu
44	10	100.0	318	20	T33006	T33006 EST56279 Hu
45	10	100.0	823	51	AT34960	AT34960 at05c08.x

ALIGNMENTS

RESULT 1
LOCUS D20150 336 bp mRNA
DEFINITION HUMGS01123 Human primate cell cycle Hom sapiens CDNA clone pm2161 3',
ACCESSION D20150
NID 9501247
VERSION D20150.1 GI:501247

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Okubo, K., Fukushima, A., Yoshii, J., Niyama, T., Kojima, Y.,
Yoshinari, H., Arimoto, J. and Matsubara, K.
Gene expression of human promyelocytic cell line HL60 before and after induction of differentiation. A new application of 3'directed cDNA sequencing
Unpublished (1993)
JOURNAL COMMENT

FEATURES
source
1. 336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="pm2161"
/note="Female, adult, cell_line = HL60, cell_type = promyelocyte."
BASE COUNT 65 a 87 c 87 t 13 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 83 ACCCAGACAG 92

RESULT 2
LOCUS D22931
DEFINITION R1C1816a Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D22931
NID 9426856
VERSION D22931.1 GI:426856
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

FEATURES
source
1. 434
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_lib="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 104 ACCCAGACAG 95

RESULT 4
LOCUS D24802/c
DEFINITION R1C2R578A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24802
NID 9428650

BASE COUNT 50 a 89 c 112 g 37 t 7 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 295;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 28 ACCCAGACAG 37

RESULT 3
LOCUS D24421/c
DEFINITION R1C18169a Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24421
NID 9428273
VERSION D24421.1 GI:428273
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 83 ACCCAGACAG 92

RESULT 2
LOCUS D22931
DEFINITION R1C1816a Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D22931
NID 9426856
VERSION D22931.1 GI:426856
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

FEATURES
source
1. 434
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_lib="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 104 ACCCAGACAG 95

RESULT 4
LOCUS D24802/c
DEFINITION R1C2R578A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24802
NID 9428650

BASE COUNT 50 a 89 c 112 g 37 t 7 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 295;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 28 ACCCAGACAG 37

RESULT 3
LOCUS D24421/c
DEFINITION R1C18169a Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24421
NID 9428273
VERSION D24421.1 GI:428273
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 83 ACCCAGACAG 92

RESULT 2
LOCUS D22931
DEFINITION R1C1816a Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D22931
NID 9426856
VERSION D22931.1 GI:426856
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

FEATURES
source
1. 434
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_lib="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 104 ACCCAGACAG 95

RESULT 4
LOCUS D24802/c
DEFINITION R1C2R578A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24802
NID 9428650

BASE COUNT 50 a 89 c 112 g 37 t 7 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 295;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 28 ACCCAGACAG 37

RESULT 3
LOCUS D24421/c
DEFINITION R1C18169a Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24421
NID 9428273
VERSION D24421.1 GI:428273
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 83 ACCCAGACAG 92

RESULT 2
LOCUS D22931
DEFINITION R1C1816a Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D22931
NID 9426856
VERSION D22931.1 GI:426856
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

FEATURES
source
1. 434
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_lib="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 104 ACCCAGACAG 95

RESULT 4
LOCUS D24802/c
DEFINITION R1C2R578A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24802
NID 9428650

BASE COUNT 50 a 89 c 112 g 37 t 7 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 295;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCCAGACAG 10
Db 28 ACCCAGACAG 37

RESULT 3
LOCUS D24421/c
DEFINITION R1C18169a Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION D24421
NID 9428273
VERSION D24421.1 GI:428273
KEYWORDS EST.
ORIGIN Oryza sativa.
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 136 a 78 c 105 g 113 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 336;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10


```

VERSION D24802.1 GI:428650
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Liliopsida; Poales;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 351)
AUTHORS Minobe,Y. and Sasaki,T.
TITLE Rice cDNA from root
JOURNAL Unpublished (1995)
COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

FEATURES
source
1..351
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone_id="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 112 a 65 c 85 g 87 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 351;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
|||||
Db 104 ACCCACACAG 95

RESULT 5
D36166/c 360 bp mRNA EST 08-AUG-1994
LOCUS CEK029H4F yuji_kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone YK29H4 5', mRNA sequence.
GSS 36166
VERSION 9528104
KEYWORDS GI:528104
SOURCE EST.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsui,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
LOCATION/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"

FEATURES
source

VERSION D24802.1 GI:428650
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Liliopsida; Poales;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 344)
AUTHORS Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
PROJECT "RGP".

FEATURES
source
1..344
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone_id="R"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 73 a 77 c 91 g 103 t
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 344;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCCACACAG 10
|||||
Db 271 ACCCACACAG 262

RESULT 7
F01745/c 260 bp mRNA EST 02-FEB-1995
LOCUS HSC0C082 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION c-0c08 3', mRNA sequence.
GSS F01745
VERSION 9645302
KEYWORDS GI:645302
SOURCE EST.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Mammalia; Primates; Hominidae; Homo.
1 (bases 1 to 260)
AUTHORS
TITLE
JOURNAL
COMMENT

```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani, K., Kachich, C., and Tessier, A.
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read. removed at sequence 5'end
 Genexpress_library_id: C; Genexpress_sequence_id: alc-0cd08
 Seq primer: (-21)M13 universal.
FEATURES Location/Qualifiers
 1..260
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="17 (15)"
 /clone="C-0cd08"
 /clone_lib="normalized infant brain CDNA"
 /sex="female"
 /tissue="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: latmid BA; Site: 1: HindIII; Site: 2: NotI; sex=Female; dev_stage=3 months old; Isolate=muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the latmid BA vector. Clone library from B.Scoates, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"
BASE COUNT 67 a 53 c 59 g 81 t
ORIGIN
 Query Match 100.0%; Score 10; DB 20; Length 260;
 Best Local Similarity 100.0%; Pred. No. 7.7e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
LOCUS M78221/c 409 bp mRNA EST 26-MAY-1992
DEFINITION EST01814 Subtracted Hippocampus, Striatum (cat. #936205) Homo sapiens cDNA clone HHCPS22 similar to Myelin basic protein, mRNA sequence.
ACCESSION M78221
VERSION 9273958
KEYWORDS M78221.1 GI:273958
SOURCE EST.
ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Uterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.

TITLE Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 632-634 (1992)
MEDLINE 92168112
COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@igf.org
 Seq primer: M13 Forward.
FEATURES Location/Qualifiers
 1..409
 /organism="Homo sapiens"
 /db_xref="ATCC (1host):78662"
 /db_xref="GDB:D051215E"
 /db_xref="taxon:9606"
 /clone="HHCPS22"
 /clone_lib="Subtracted Hippocampus, Striatum (cat. #936205)"
 /note="Vector: lambdaZAP-II; The hippocampus library (#4 above) was subtracted with a fibroblast cell line cDNA library (Striatum cat. #936209; M13 lung fibroblast cell line: oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average insert size.) by the method of Sive & St. John (Nucl. Acids Res. 16:10937, 1988)."
BASE COUNT 101 a 87 c 100 g 120 t 1 others
ORIGIN
 Query Match 100.0%; Score 10; DB 20; Length 409;
 Best Local Similarity 100.0%; Pred. No. 8.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
LOCUS M78683/c 386 bp mRNA EST 11-JAN-1995
DEFINITION EST00831 Hippocampus, Striatum (cat. #936205) Homo sapiens cDNA clone HHCPS22, mRNA sequence.
ACCESSION M78683
VERSION 9272998
KEYWORDS M78683.1 GI:272998
SOURCE EST.
ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Uterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
TITLE Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 632-634 (1992)
MEDLINE 92168112
REFERENCE 2. (bases 1 to 386)
AUTHORS Durbin, A.S.
TITLE EST mapping data
JOURNAL Unpublished (1996)
COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@igf.org
 Seq primer: M13 Reverse.


```

FEATURES
  source
    Location/Qualifiers
      1..386
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):78089"
        /db_xref="GDB:DL6S442E"
        /map="16 p13.13"
        /clone="HHCPC05"
        /note="Vector: lambdaZAP-II; Female, 2 years; oligo-dT +
        random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb
        average insert size."
BASE COUNT      83 a      100 c      98 g      103 t      2 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 386;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACCCACACAG 10
|||||
191 ACCCACACAG 182

RESULT 10
M79123      423 bp      mRNA      EST      26-MAY-1992
LOCUS      EST01271 Subtracted Hippocampus, Strata gene (cat. #936205) Homo
DEFINITION  sapiens cDNA clone HHCPC02, mRNA sequence.
ACCESSION   M79123
NID         927436
VERSION     M79123.1 GI:273436
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
COMMENT     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 423)
AUTHORS    Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
TITLE      Sequence identification of 2,375 human brain genes
JOURNAL     Nature 355, 632-634 (1992)
MEDLINE     92168112
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seg primer: M13 Forward.
FEATURES
  source
    Location/Qualifiers
      1..423
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):78667"
        /db_xref="GDB:DOS2091E"
        /db_xref="taxon:9606"
        /clone="HHCPC02"
        /clone.lib="Subtracted Hippocampus, Strata gene (cat.
        #936205)"
        /note="Vector: lambdaZAP-II; The hippocampus library (#4
        above) was subtracted with a fibroblast cell line cDNA
        library (Stratagene cat. #936209; W38 lung fibroblast
        cell line; oligo-dT + random primed cDNA synthesis;
        lambdaZAP-II vector, 1.0kb average insert size.) by the
        method of Sive & St. John (Nucl. Acids Res. 16:10937,
        1988)."
BASE COUNT      145 a      86 c      67 g      123 t      2 others
ORIGIN

```

```

Query Match      100.0%; Score 10; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCCACACAG 10
|||||
Db      74 ACCCACACAG 83

RESULT 11
M79888      312 bp      mRNA      EST      30-JUN-1992
LOCUS      M79888/c
DEFINITION  W69700425 Mixed stage, Strata gene (cat. #937006) Caenorhabditis
elegans cDNA clone CEMS705 similar to Actin, mRNA sequence.
ACCESSION   M79888
NID         9271907
VERSION     M79888.1 GI:271907
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
COMMENT     Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 312)
AUTHORS    McComble,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,
            Utterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and
            Fields,C.
TITLE      Caenorhabditis elegans expressed sequence tags identify gene
JOURNAL     families and potential disease gene homologues
MEDLINE     Nature Genet. 1, 124-131 (1992)
          93250983
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seg primer: M13 Forward.
FEATURES
  source
    Location/Qualifiers
      1..312
        /organism="Caenorhabditis elegans"
        /db_xref="taxon:6239"
        /clone="CEMS705"
        /clone.lib="Mixed stage, Strata gene (cat. #937006)"
        /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA
        library; Stratagene catalog #937006. The library is oligo
        dT primed and directionally cloned in the Uni-ZAP XR
        vector."
BASE COUNT      78 a      64 c      103 g      64 t      3 others
ORIGIN
Query Match      100.0%; Score 10; DB 20; Length 312;
Best Local Similarity 100.0%; Pred. No. 7.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCCACACAG 10
|||||
Db      228 ACCCACACAG 219

RESULT 12
M85714      292 bp      mRNA      EST      26-MAY-1992
LOCUS      M85714/c
DEFINITION  EST02223 Fetal brain, Strata gene (cat.#936206) Homo sapiens cDNA
clone HFCM42, mRNA sequence.
ACCESSION   M85714
NID         9274361
VERSION     M85714.1 GI:274361
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```


REFERENCE 1 (bases 1 to 292)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT

CONTACT: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 forward.
 Location/Qualifiers
 1..292
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):81291"
 /db_xref="GDB:DS407E"
 /db_xref="taxon:9606"
 /clone_lib="Fetal brain, Stratagene (cat#936206)"
 /note="Vector: LambdaZAP-II: 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."
 BASE COUNT 73 a 57 c 72 g 89 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 292;
 Best Local Similarity 100.0%; Pred. No. 7.8e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACCCACACAG 10
 |||||
 Db 82 ACCCACACAG 73

RESULT 13
 M85791 323 bp mRNA EST 26-MAY-1992
 LOCUS EST02312 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
 DEFINITION clone HFBNCN71, mRNA sequence.
 ACCESSION M85791
 9274440
 M85791.1 GI:274440
 EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 323)
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT

CONTACT: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 forward.
 Location/Qualifiers
 1..323
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):81368"

/db_xref="GDB:DS482E"
 /db_xref="taxon:9606"
 /clone="HFBNCN71"
 /clone_lib="Fetal brain, Stratagene (cat#936206)"
 /note="Vector: LambdaZAP-II: 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."
 BASE COUNT 74 a 84 c 57 g 102 t 6 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 323;
 Best Local Similarity 100.0%; Pred. No. 7.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACCCACACAG 10
 |||||
 Db 25 ACCCACACAG 34

RESULT 14
 M88939 519 bp mRNA EST 02-DEC-1992
 LOCUS CEL12B2 Chris Martin sorted cDNA library Caenorhabditis elegans
 DEFINITION cDNA clone cm12b2 5' similar to actin - C. elegans, mRNA sequence.
 ACCESSION M88939
 9275440
 M88939.1 GI:275440
 EST.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditida;
 Rhabditina; Rhabditidae; Rhabditidae; Peloderiinae; Caenorhabditis.
 1 (bases 1 to 519)
 AUTHORS Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A.,
 Hillier, L., Durbin, R.K., Green, P., Showkeen, R., Halloran, N.,
 Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J.
 and Sulston, J.
 TITLE A survey of expressed genes in Caenorhabditis elegans
 JOURNAL Nature Genet. 1, 114-123 (1992)
 MEDLINE 93250982
 COMMENT

CONTACT: Waterston R.H. (USA) and Sulston J.E. (UK)
 (USA) Dept. of Genetics or (UK)
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of
 Molecular Biology
 Box 8232, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills
 Road, Cambridge CB2 2QH, UK
 Tel: (USA) (314)3627072 or (UK) (0223)248011
 Fax: (USA) (314)3624137 or (UK) (0223)402008
 Email: rwater@genet.wustl.edu or jes@nc-cimba.cambridge.ac.uk
 Single read.
 Location/Qualifiers
 1..519
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="cm12b2"
 /clone_lib="Chris Martin sorted cDNA library"
 /lab_host="MC1061"
 /note="Vector: lambdaZAP II. Partially normalized by
 hemaphysalite cDNA library. Partially normalized by
 successively picking groups of clones that didn't
 hybridize to previously picked clones. Vector: lambdaZAP II
 SHLX2 (Lipschitz, D.H. et al., Gene 88:25-36 (1990)) Host:
 MC1061"

BASE COUNT 121 a 162 c 106 g 128 t 2 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 519;
 Best Local Similarity 100.0%; Pred. No. 8.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAG 10
 |||||
 Db 156 ACCCAGACAG 165

RESULT 15

T03144

LOCUS T03144 376 bp mRNA EST 14-FEB-1997
 DEFINITION FB25F5 Fetal brain, Strata gene Homo sapiens cDNA clone FB25F5
 3' end, mRNA sequence.

ACCESSION

T03144

NID

9314385

VERSION

T03144.1 GI:314385

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 376)

AUTHORS

Khan, A.S., Wilcox, A.S., Polymenopoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K., and Sikela, J.M.

TITLE

Single pass sequencing and physical and genetic mapping of human

JOURNAL

CDNAs

MEDLINE

Nature Genet. 2, 180-185 (1992)

COMMENT

94258200

On Sep 21, 1992 this sequence version replaced gi:279283.

Contact: Sikela JM

Department of Pharmacology

University of Colorado Health Sciences Center

Box C236, 4200 E. 9th Ave, Denver CO 80262-0236

Tel: 3032708637

Fax: 3032707097

Email: mikki@italy.uchsc.edu

Seq primer: -21M13 universal.

Location/Qualifiers

1. 376

/organism="Homo sapiens"

/db_xref="ATCC (inhost):86789"

/db_xref="GDB:D082912E"

/db_xref="taxon:9606"

/clone="FB25F5"

/clone_1b="Fetal brain, Strata gene"

/note="Vector: Lambda ZAP II; The FB 11 library (catalog #937201, Strata gene) was constructed by directional cloning and oligo d(T)-priming in lambda ZAP II phage utilizing mRNA from a 17 week human fetus (total brain)."

-FE COUNT

101 a

ORIGIN

90 c 109 t 3 others

Query Match

100.0%: Score 10; DB 20; Length 376;

Best local Similarity 100.0%: Pred. No. 8e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAG 10

|||||

Db 21 ACCCAGACAG 30

Search completed: September 17, 1999, 21:28:01
 Job time: 14265 sec

DR P-PSDB; R03724.

PT Human T-cell leukemia virus type-1 gag antigens -
 PT Produced in large amounts by expression in yeast, used in
 PT diagnostic kits or as HTLV-1 vaccine.
 PS Disclosure: fig. 3; 15pp; English.
 CC The gag (group specific antigen) proteins can be mass-
 CC produced using Saccharomyces cerevisiae transformants. The
 CC HTLV-1 gag gene or part of it can be prepared from an HTLV-1
 CC pro virus using appropriate restriction enzymes. A large
 CC amount of gag gene or part of it can be prepared by
 CC amplifying in Escherichia coli a plasmid subcloned with the
 CC HTLV-1 gag gene. DNA that encodes the gag protein is then
 CC inserted into an expression vector. The proteins obtained
 CC can be used in a diagnostic kit or as a HTLV-1 vaccine. See
 CC also R03724.
 SQ Sequence 1290 BP; 323 A; 465 C; 259 G; 243 T;

Query Match 100.0%; Score 10; DB 1; Length 1290;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LY 1 CCCTATGGAG 10
 DB 1278 CCCTATGGAG 1269

RESULT 3
 ID 003601 standard; DNA; 1891 BP.

AC 003601; 02-AUG-1990 (first entry)
 DE Gene encoding yeast aminopeptidase I.
 KW Aminopeptidase; AP1; AP2; APX; vacuolar localisation; ds.
 OS Saccharomyces cerevisiae.
 FH Key location/Qualifiers
 FT cds 166..487
 FT transit_peptide 166..300
 FT /tag= a
 FT /label= vacuolar localisation presequence.
 FT mat_peptide 301..487
 FT /tag= c
 PN EP-359164-A.
 PD 21-MAR-1990.
 PE 9-SEP-1989; 116734.
 PF 13-SEP-1988; US-243733.
 PR 14-DEC-1988; US-284244.
 (GEO-) Gen Hospital Corp.
 Smith JA, Chang Y;
 WPI: 90-084991/12.
 P-PSDB: R05535.

DR Substrate-specific amino:peptidase enzymes AP1, AP2 and APX -
 PT used to remove amino-terminal amino acid residues from
 PT peptide(s) and protein(s).
 PS Disclosure: 24pp; English.
 CC Aminopeptidases are useful in extending the stability, solubility and
 CC biological activity and half-life, or in decreasing the immunogenicity.
 CC The vacuolar location presequence encoded by the gene is useful in
 CC directing linked polypeptides through the vacuolar membrane.
 SQ Sequence 1891 BP; 606 A; 387 C; 388 G; 510 T;

Query Match 100.0%; Score 10; DB 1; Length 1891;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGGAG 10
 DB 640 CCCTATGGAG 649

RESULT 4
 ID N90123

ID N90123 standard; DNA; 5524 BP.
 AC N90123;
 DT 1-NOV-1989 (first entry)
 DE DNA encoding human common acute lymphoblastic leukaemia antigen
 KW human common acute lymphoblastic leukaemia antigen;
 OS analgesic; inhibits leukaemia; endopeptidase.
 FH Homo sapiens (Human)

FT cds 12
 FT /tag= a
 PN WO8905353-A.
 PD 15-JUN-1989.
 PE 01-DEC-1988; U04280.
 PR 04-DEC-1987; US-247915.
 PA (DANA) Dana Farber Cancer Inst.
 PI Reinherz EL, Shipp MA, Richardson NE, Ritz J, Sayre PH;
 DR WPI: 89-192699/26.
 P-PSDB: P90393.
 PT DNA encoding human common acute lymphoblastic leukaemia antigen
 PT - used for obtaining pure protein for diagnosis and
 PT treatment of medical conditions.
 PS Claim 4; fig 3; 73pp; English.
 CC DNA encoding human common acute lymphoblastic leukaemia
 CC antigen (CALA) (see corresp. P90393), and its fragments. Binds to
 CC anti-CALA antibody, and is used for diagnosis, treatment, to
 CC design a drug for use as a general analgesic. The DNA has functional
 CC endopeptidase activity.
 SQ Sequence 5524 BP; 1777 A; 1039 C; 1066 G; 1642 T;

Query Match 100.0%; Score 10; DB 1; Length 5524;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGGAG 10
 DB 4546 CCCTATGGAG 4555

RESULT 5
 ID N80297/c
 AC N80297; standard; RNA; 4992 BP.

DT 16-SEP-1990 (first entry)
 DE Partial nucleic acid sequence of the Japanese encephalitis virus (JEV)
 DE genome encoding C, M, E and NS proteins
 KW Japanese encephalitis virus (JEV); vaccine; flavivirus immunity;
 KW C protein; M protein; E protein; NS protein.
 OS Japanese encephalitis virus.
 FH Key location/Qualifiers
 FT mat_peptide 1..159
 FT /tag= a
 FT /product= C protein
 FT cds 160..441
 FT /tag= b
 FT /product= pre M protein
 FT mat_peptide 442..666
 FT /tag= c
 FT /product= M protein
 FT mat_peptide 667..2166
 FT /tag= d
 FT /product= E protein
 FT mat_peptide 2167..3408
 FT /tag= e
 FT /product= NS1
 FT mat_peptide 3409..3903
 FT /tag= f
 FT /product= ns2a
 FT mat_peptide 3904..4296
 FT /tag= g
 FT /product= ns2b
 FT mat_peptide 4297..4992
 FT /tag= h


```

FT      /product-NS3
PN      MO8803032-A.
PD      05-MAY-1988.
PF      21-OCT-1987; 002763.
PR      27-OCT-1986; US-923907.
PA      (FOUR/) Fournier MJ.
PI      MOAda PC, Mason PM.
DR      WPI; 88-133144/19.
DR      P-PSDB; P80264.
PT      Nucleic acid of Japanese encephalitis virus -
PT      used for producing polypeptide(s) and in the diagnosis of and
PS      immunisation against the virus
PS      Claim 4: Fig 5/1-5/5: 52pp; English.
CC      Substantially purified nucleic acid comprising at least a 10 bp sequence
CC      of DNA corresponding to the nucleic acid sequence of JEV but not to the
CC      nucleic acid sequence of yellow fever virus wherein the 10 bp sequence
CC      is included within the sequence shown in n80297. Pref. the nucleic acid
CC      sequences do not correspond to those found in yellow fever virus, West
CC      Nile virus, Murray valley fever virus, dengue virus or St. Louis
CC      encephalitis virus. The nucleic acid may be obtd. from the virus and
CC      inserted into a vector. Specific probes can be derived from this
CC      nucleic acid. Also claimed are a pure polypeptide synthesized from the
CC      purified nucleic acid and antibodies produced to the polypeptide. The
CC      polypeptides and nucleic acid prods. are used in the diagnosis of and
CC      immunisation against JEV. Inoculation of animals with synthetic viral
CC      protein immunogens or with active vectors encoding such immunogens,
CC      elicits protective antiviral antibodies. The vaccination induces
CC      immunity to the disease caused by other flaviviruses.
SQ      Sequence 4992 BP; 1332 A; 1166 C; 1390 G; 1100 U; 4 Others;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 4992;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCCTATGAG 10
DB      1980 CCCTATGAG 1971

RESULT 6
N81434/C
ID      N81434 standard; DNA; 214 BP.
AC      N81434;
DE      11-DEC-1990 (first entry)
DE      Autonomously replicating sequence.
DE      Autonomously replicating; insulin; growth hormone; Interferons; TNF;
DE      Interleukins; lymphokines; enzymes; ss.
PI      Key
PI      Location/Qualifiers
FT      misc_rna
FT      80..177
FT      /tag= a
FT      /label=active fragment
FT      80..158
FT      /tag= b
FT      /label=active fragment
FT      80..105
FT      /tag= c
FT      /label=active fragment
FT      130..158
FT      /tag= d
FT      /label=active fragment
FT      150..177
FT      /tag= e
FT      /label=active fragment
FT      130..177
FT      /tag= f
FT      /label=active fragment
PN      EP-254315-A.
PD      27-JAN-1988.
PF      23-JUL-1987; 110696.
PR      24-JUL-1986; JP-174036.
PR      26-SEP-1986; JP-227455.
PA      (DAUC ) DAICHI SEIYAKU KK.

```

```

PI      Ariga H;
DR      WPI; 88-022976/04.
PT      Plasmids contg. autonomously replicating sequence DNA - useful
PT      for producing proteins
PS      Claim 9: page 10; 17pp; English.
CC      This sequence, and active fragments, is contained in an expression
CC      plasmid alongside a promoter and a gene encoding a protein of int-
CC      erest. The plasmid is propagated in host cells for the prodn. of
CC      e.g. insulin, growth hormone, TNF and lymphokines. This autonom-
CC      ously replicating sequence has an affinity for a DNA-binding prot-
CC      ein and may be recovered by binding a mammalian cell-derived DNA
CC      fragment/myc protein prod.
CC      See also N81433.
SQ      Sequence 214 BP; 37 A; 46 C; 45 G; 86 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 214;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCCTATGAG 10
DB      100 CCCTATGAG 91

RESULT 7
006691
ID      006691 standard; DNA; 2175 BP.
AC      006691;
DE      27-FEB-1991 (first entry)
DE      CDX, a MILA involved in ELAM1-mediated adhesion, from PCDM8 clone 7.2.
DE      Endothelial cell-leucocyte adhesion molecule 1; ELAM1; CDX;
DE      KW molecule involved in leucocyte adhesion; MILA; inflammation; ss.
OS      Homo sapiens.
PI      Key
PI      Location/Qualifiers
FT      cds
FT      66..1280
FT      /tag= a
FT      /product= CDX
FT      /note= "a MILA "
PN      M09013300-A.
PD      15-NOV-1990.
PF      27-APR-1990; 002357.
PR      28-APR-1989; US-345151.
PR      01-JUN-1989; US-359516.
PR      18-DEC-1989; US-452675.
PA      (BIOG-) BIOGEN INC.
PI      Hession C, Lobb RR, Goeltz SE, Born I, Benjamin CD;
PI      Rose MD;
PI      WPI; 90-361248/48.
DR      P-PSDB; R08119.
PT      Endothelial cell adhesion mol. - MILAs and DNA encoding them and
PT      inhibition-detection of binding of leukocytes to endothelial
PT      cells
PS      Disclosure: Fig 9 (A-C); 136pp; English.
CC      The sequence was obtained from CDX pCDM8 clone 7.2 and from a
CC      portion of the 7.2 insert subcloned into the sequencing vector pNN11.
CC      CDX is a MILA, involved in ELAM1-mediated adhesion and is probably
CC      the (or an) ELAM1 ligand. CDX is expressed in melanomas.
CC      See also 006686-91.
SQ      Sequence 2175 BP; 403 A; 643 C; 652 G; 477 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 2175;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CCCTATGAG 10
DB      2039 CCCTATGAG 2048

RESULT 8
011140

```



```

PD 25-JUL-1991.
PF 08-JAN-1991; U00145.
PR 08-JAN-1990; US-462158.
PA (RESE.) RES CORP TECHN INC.
PI Vansnick J, Ytteenhove C, Simpson RJ;
DR WPI: 91-238026/32.
DR P-PSDB: R13218.
PT Nucleic acid encoding for P40 T-cell growth factor - for patients
PT with AIDS or compromised immune systems, also allows increased
PT prodn. of other cytokine(s).
PS Claim 1; Fig 15; 122pp; English.
CC The sequence was obtd. from a clone, lambda S40.1a, isolated from
CC a genomic library using a murine P40 cDNA clone as a probe. The
CC sequence, or the cDNA sequence obtd. using the sequence, can be
CC inserted into a vector for expression of P40 in a host organism.
CC The protein is useful for stimulating the proliferation of certain
CC subsets of T helper cells e.g. in AIDS patients or immune
CC compromised patients.
CC See also Q12759 and Q13660-Q13663.
SQ Sequence 3809 BP; 1013 A; 903 C; 869 G; 1024 T;

Query Match 100.0%; Score 10; DB 1; Length 3809;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 CCCATGAG 10
|||||||
DB 3040 CCCATGAG 3031

RESULT 12
ID Q14382
AC Q14382; standard; DNA; 2175 BP.
DT 13-FEB-1992 (first entry)
DE Clone 7.2 encoding 1,3-fucosyl transferase.
KM CDX; cell adhesion; ELAM1; ss.
OS Homo sapiens.
PF Key
FT Location/Qualifiers
FT cds 66..1283
FT /tag= a
PN WO9116900-A.
PD 14-NOV-1991.
PF 26-OCT-1990; U06198.
PR 27-APR-1990; WO-U02357.
PR 26-OCT-1990; WO-U06198.
PA (BIOJ.) BIOGEN INC.
PI Goelz SE, Hession CA;
DR WPI: 91-353507/48.
DR P-PSDB: R14404.
PT DNA sequences encoding 1,3-fucosyl transferase - used to develop
PT antiinflammatory therapy by inhibition of linking activity.
PS Claim 1; Fig 1; 38pp; English.
CC The clone was obtd. from CDX PCDM8 clone 7.2 and from a portion of
CC the 7.2 insert subcloned into the sequencing vector pNN1 to
CC prepare pSQ219. The sequence encodes protein 7.2, a 1,3-fucosyl
CC transferase and which binds to ELAM1. The portion of the sequence
CC from nucleotide 9-2162 is identical to the portion of clone 1 (Q14383)
CC from nucleotide 493-2645. The two inserts may represent different
CC transcripts from the same DNA segment. The protein may be useful in
CC the development of anti-inflammatory or other therapies.
CC See also Q14383.
SQ Sequence 2175 BP; 403 A; 647 C; 648 G; 477 T;

Query Match 100.0%; Score 10; DB 1; Length 2175;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 CCCATGAG 10
|||||||

```


Db 2039 CCCTATGAG 2048

RESULT 13

ID 014383 standard; DNA; 2861 BP.

AC 014383;

DE 13-FEB-1992 (first entry)

DE Clone 1 encoding 1,3-fucosyl transferase.

KM CDX; cell adhesion; ELAM1; ss.

OS Homo sapiens.

FH key Location/Qualifiers

FT cds 74..1766

FT /tag- a

PN MO9116900-A.

PD 14-NOV-1991.

PF 26-OCT-1990; WO-06198.

PR 27-APR-1990; WO-002357.

PR 26-OCT-1990; WO-006198.

PR (BioJ) BIOGEN INC.

PR Goelz SE, Hession CA;

PR P-PSDB: R14405

PT DNA sequences encoding 1,3-fucosyl transferase - used to develop

PT antiinflammatory therapy by inhibition of linking activity.

PS Claim 1; Fig 2; 38pp; English.

CC The clone was obt. from CDX PCDM8 clone 1. The sequence encodes

CC protein 1, a 1,3-fucosyl transferase a surface glycoprotein which

CC is recognised by anti-CDX antibodies and which binds to ELAM1. The

CC portion of the sequence from nucleotide 492-2645 is identical to the

CC portion of clone 7.2 (Q14382) from nucleotide 9-2162. The two inserts

CC may represent different transcripts from the same DNA segment. The

CC protein may be useful in the development of anti-inflammatory or other

CC therapies.

CC See also Q14382.

SQ Sequence 2861 BP; 541 A; 840 C; 881 G; 599 T;

Query Match 100.0%; Score 10; DB 1; Length 2861;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 2522 CCCTATGAG 2531

3UTR 14

N50275 standard; DNA; 1530 BP.

AC N50275;

DE 30-NOV-1991 (first entry)

DE Sequence from PATK105 contg. the p24 coding region and flanked by

DE gag gene and 5' LTR.

KM Antigen; diagnosis; assay; blood transfusion; ss.

OS Adult T-cell leukaemia virus (ATLV).

FH key Location/Qualifiers

FT ltr 1..35

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- p24

FT /tag- d

FT /product- gag

FT /tag- e

FT /tag- f

FT /tag- g

FT /tag- h

FT /tag- i

FT /tag- j

FT /tag- k

FT /tag- l

FT /tag- m

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the AUG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pIR33 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Registration Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DE 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

FT /tag- q

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the AUG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pIR33 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Registration Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DE 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

FT /tag- q

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the AUG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pIR33 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Registration Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DE 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

FT /tag- g

FT /product- px-III

FT /tag- h

FT /product- px-IV

FT /tag- i

FT /product- px-V

FT /tag- j

FT /product- px-VI

FT /tag- k

FT /product- px-VII

FT /tag- l

FT /product- px-VIII

FT /tag- m

FT /product- px-IX

FT /tag- n

FT /product- px-X

FT /tag- o

FT /product- px-XI

FT /tag- p

FT /product- px-XII

FT /tag- q

PI Taniguchi T, Yoshida M, Sugano H, Sekine S;

DR WPI: 85-106374/18.

DR P-PSDB: P50232.

PT New adult T-cell leukaemia virus antigen peptide - used in the

PT detection of T-cell leukaemia carriers in blood donors

PS Example; Table 1, Page 15-16; 37pp; English.

CC PATK105 contains a sequence extending from the 18th base pair

CC upstream from the AUG codon to the 139th base downstream from the

CC terminator codon of the ATLV gene. The example concerns the cloning

CC of the gag gene of PATK105 into pIR33 vector, which produced

CC pTAC424A which codes for a polypeptide consisting of a total of 317

CC AAs, wherein 17 AAs and 85 AAs derived from the gag gene are

CC respectively attached to the N-terminal and C-terminal of p24.

CC Escherichia coli K-12, HB101 strain which contains plasmid pTAC424A

CC has been deposited with the Registration Research Institute as

CC Escherichia coli ETAC424A, FERM BP-341.

SQ Sequence 1350 BP; 386 A; 546 C; 306 G; 292 T;

Query Match 100.0%; Score 10; DB 1; Length 1350;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTATGAG 10

Db 1359 CCCTATGAG 1350

RESULT 15

ID N40080 standard; cDNA; 9047 BP.

AC N40080;

DE 04-FEB-1992 (first entry)

DE Sequence complementary to the genome of adult T-cell leukaemia virus

DE (ATLV) ATK-1.

KM Diagnosis; therapy; prevention; leukaemia; antigen; ss.

OS Adult T-cell leukaemia virus.

FH key Location/Qualifiers

FT 5'utr 1..755

FT /tag- a

FT /tag- b

FT /product- gag

FT /tag- c

FT /product- pol

FT /tag- d

FT /product- env

FT /tag- e

FT /product- px-I

FT /tag- f

FT /product- px-II

PI Yoshida M, Sugano H;
DR WPI; 84-172336/28.
PT Viral genomic DNA complementary to RNA of human leukaemia virus -
PT useful in recombinant DNA producing therapeutic and diagnostic
PT proteins
PS Claim 4; Table 1, Page 11-15; 23pp; English.
CC Diagnosis of human leukaemia and/or lymphoma and virus infection is
CC made with part or all of the recombinant DNA. Virus antigenic
CC proteins can be produced. These peptides and proteins, and
CC antibodies against them, are useful for the diagnosis, therapy and
CC prevention of human leukaemia.
SQ Sequence 9047 BP; 2087 A; 3164 C; 1713 G; 2083 T;

Query Match 100.0%; Score 10; DB 1; Length 9047;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCCTATGGAG 10
|||||||
2079 CCCTATGGAG 2070

Search completed: September 18, 1999, 00:33:35
Job time: 18975 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:01 : Search time 2825.05 Seconds

(without alignments)
6.962 Million cell updates/sec

Title: US-08-956-518a-90
Perfect score: 10
Sequence: 1 CCTATGAG 10

Scoring table: IDENTITY_NUC

arched: 2546578 seqs, 98626752 residues

-cbase :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	257	20	D19706	D19706 HUMG00662
2	10	100.0	360	20	D35906	D35906 CEIK02685F
3	10	100.0	360	20	D35977	D35977 CEIK02777F
4	10	100.0	298	20	D39999	D39999 RICS1711A R
5	10	100.0	188	20	F01533	F01533 HSC05H042 n
6	10	100.0	312	20	F03038	F03038 HSC01KD092 n
7	10	100.0	349	20	M85318	M85318 EST01434 Fe
8	10	100.0	335	20	T01239	T01239 WST01960 E
9	10	100.0	667	20	T03839	T03839 GT0024 Gene
10	10	100.0	422	20	T03843	T03843 GT0131 Gene
11	10	100.0	236	20	T03944	T03944 GT1159 Gene
12	10	100.0	400	20	T04417	T04417 464 Lambda-
13	10	100.0	413	20	T04785	T04785 833 Lambda-
14	10	100.0	357	20	T06534	T06534 EST04423 Fe
15	10	100.0	453	20	T13875	T13875 2040 Lambda
16	10	100.0	389	20	T18860	T18860 B090205 Tes
17	10	100.0	149	20	T25900	T25900 ATH014 HPCD
18	10	100.0	339	20	T26587	T26587 AB327H2F In
19	10	100.0	264	20	T27698	T27698 EST12669 Hu
20	10	100.0	223	20	T29507	T29507 EST82429 Hu
21	10	100.0	210	20	T31409	T31409 EST31977 Hu
22	10	100.0	306	20	T31520	T31520 EST34096 Hu
23	10	100.0	329	20	T31768	T31768 EST38526 Hu
24	10	100.0	348	20	T32482	T32482 EST49297 Hu
25	10	100.0	297	20	T32537	T32537 EST50461 Hu
26	10	100.0	307	20	T33954	T33954 EST60289 Hu
27	10	100.0	381	20	T34070	T34070 EST62150 Hu
28	10	100.0	319	20	T35772	T35772 EST91174 Hu
29	10	100.0	482	20	T50468	T50468 YB30G01.s1
30	10	100.0	372	20	T52181	T52181 YB2811.s1
31	10	100.0	430	20	T54450	T54450 YB06H04.r2
32	10	100.0	237	20	Z22575	Z22575 m142/1AB In
33	10	100.0	320	20	Z23513	Z23513 ATTS1026 Gr
34	10	100.0	168	20	Z31029	Z31029 NMTEST128 M
35	10	100.0	412	20	Z37642	Z37642 ATTS4107 Ve
36	10	100.0	317	20	Z40389	Z40389 HSC21B082 n
37	10	100.0	337	20	Z42738	Z42738 HSC05E101 n
38	10	100.0	330	20	Z42925	Z42925 HSC07A121 n
39	10	100.0	282	20	Z44843	Z44843 HSC2AG121 n
40	10	100.0	206	20	Z45940	Z45940 HSC2XG121 n
41	10	100.0	308	20	Z46007	Z46007 HSC06G021 n
42	10	100.0	258	21	F05474	F05474 HSC06C011 n
43	10	100.0	313	21	F05489	F05489 HSC06F091 n
44	10	100.0	359	21	F06540	F06540 HSC18E071 n
45	10	100.0	590	51	A1734774	A1734774 606036C04

ALIGNMENTS

RESULT 1
D19706/c D19706 257 bp mRNA EST 30-JUL-1996
LOCUS HUMG00662 Human promyelocyte Homo sapiens cDNA clone mm08f12 3',
DEFINITION MRNA sequence.
ACCESSION D19706
NID 9500022
VERSION D19706.1 GI:500022

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Okubo, K., Fukushima, A., Yoshii, J., Niijima, T., Kojima, Y.,
1 (bases 1 to 257)
TITLE Yoshinari, H., Arimoto, J. and Matsubara, K.,
Gene expression of human promyelocytic cell line HL60 before and
after induction of differentiation. A new application of 3'directed
cDNA sequencing
JOURNAL Unpublished (1993)
COMMENT
Contact: Okubo, K., Fukushima, A., Yoshii, J., Niijima, T., Kojima, Y.,
Yoshinari, H., Arimoto, J. and Matsubara, K.,
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers
FEATURES
source
1..257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human promyelocyte"
/note="Female, adult, cell_line = HL60, cell_type =
promyelocyte."
BASE COUNT 84 a 35 c 54 g 79 t 5 others
ORIGIN
Query Match 100.0%; Score 10; DB 20; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTATGAG 10
|||||
Db 174 CCTATGAG 165
RESULT 2
D35906 360 bp mRNA EST 08-AUG-1994
LOCUS CELK0265F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone YK26h5 5', mRNA sequence.
ACCESSION D35906
NID 9527845
VERSION D35906.1 GI:527845
KEYWORDS EST.
RCE
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara, Y., Mitsuiki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male,
Location/Qualifiers

BASE COUNT 104 a 66 c 83 g 103 t 4 others
ORIGIN
Query Match 100.0%; Score 10; DB 20; Length 360;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTATGAG 10
|||||
Db 154 CCTATGAG 163
RESULT 3
D35977 360 bp mRNA EST 08-AUG-1994
LOCUS CELK027E7F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone YK27F7 5', mRNA sequence.
ACCESSION D35977
NID 9527915
VERSION D35977.1 GI:527915
KEYWORDS EST.
SOURCE
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara, Y., Mitsuiki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal."
BASE COUNT 92 a 78 c 81 g 102 t 7 others
ORIGIN
Query Match 100.0%; Score 10; DB 20; Length 360;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTATGAG 10
|||||
Db 21 CCTATGAG 30
RESULT 4
D39999 298 bp mRNA EST 11-NOV-1994
LOCUS RICS1711A Rice shoot Oryza sativa cDNA, mRNA sequence.
DEFINITION RICS1711A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION D39999
NID 9569150
VERSION D39999.1 GI:569150
KEYWORDS EST.
SOURCE
ORGANISM Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphylliphytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 298)
 Sasaki,T., Miyao,A. and Yamamoto,K.
 Rice cDNA from callus 1995
 JOURNAL
 Unpublished (1995)

COMMENT

Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp.
 Location/Qualifiers

FEATURES

1. 298

/organism="Oryza sativa"
 /strain="Nipponbare, sub_species Japonica"
 /db_xref="taxon:4530"
 /map="8"
 /clone_id="Rice shoot"
 /note="Etisolated shoot (8 days old)"

BASE COUNT 91 a 61 c 62 g 79 t 5 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 298;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGAG 10
 |||||
 Db 107 CCCTATGAG 116

RESULT 5

F01533/c 188 bp mRNA EST 01-FEB-1995
 LOCUS HSC05H042 normalized infant brain cDNA Homo sapiens cDNA clone
 DEFINITION c-05H04 3', mRNA sequence.

ACCESSION F01533

VERSION F01533.1 GI:645090

WORDS EST.

RCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 188)

REFERENCE

AUTHORS

Devignes,M.D., Duprat,S., Bouquier,C., da Silva,C.,
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Juneau,M.N., Lamy,B.,
 Sebastiao,Kabakchis,C. and Tessier,A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression

TITLE

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL 95277534
 MEDLINE On Oct 24, 1994 this sequence version replaced.

COMMENT

Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698

FEATURES

SOURCE

1. 188
 Location/Qualifiers
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read. removed at sequence 5' and
 Genexpress_library_idt: C; Genexpress_sequence_idt: alc-05h04
 Seq primer: (-21)M13.universal.
 Location/Qualifiers
 1. 188

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="c-05H04"

/clone_id="normalized infant brain cDNA"

/sex="Female"

/tissue_type="total brain"

/dev_stage="3 months old"

/note="Organ: brain; Vector: lafmd BA; Site_1: HindIII;
 Site_2: NotI; sex:female; dev_stage=3 months old;
 isolate-muscular atrophy patient; tissue_type=total
 brain; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lafmd BA vector. Clone library from B Soares, Psychiatry
 Dept. Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S in press"

BASE COUNT 59 a 48 c 30 g 47 t 4 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 188;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCTATGAG 10
 |||||
 Db 156 CCCTATGAG 147

RESULT 6

F03038 312 bp mRNA EST 02-FEB-1995
 LOCUS HSC1KD092 normalized infant brain cDNA Homo sapiens cDNA clone
 DEFINITION c-1kd09 3', mRNA sequence.

ACCESSION F03038

VERSION F03038.1 GI:646595

WORDS EST.

RCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 312)

REFERENCE

AUTHORS

Devignes,M.D., Duprat,S., Bouquier,C., da Silva,C.,
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Juneau,M.N., Lamy,B.,
 Sebastiao,Kabakchis,C. and Tessier,A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression

TITLE

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL 95277534

MEDLINE

COMMENT

Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read. removed at sequence 5'end
 Genexpress_library_idt: C; Genexpress_sequence_idt: alc-1kd09
 Seq primer: (-21)M13.universal.
 Location/Qualifiers
 1. 312

FEATURES

SOURCE

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="1"
 /clone="c-1kd09"
 /clone_id="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lafmd BA; Site_1: HindIII;
 Site_2: NotI; sex:female; dev_stage=3 months old;"

isolate-muscular atrophy patient; tissue-type-total
 brain: total mRNA was oligo-(GT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambdaB A vector. Clone library from B.Soures, Psychiatry
 Dept. Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S in press"

BASE COUNT 91 a 72 c 55 g 90 t 4 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 312;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTATGAG 10
 |||||

DB 155 CCTATGAG 146

RESULT 7
 118
 US M85318 349 bp mRNA EST 26-MAY-1992
 DEFINITION EST01434 Fetal brain, Striatum (cat#936206) Homo sapiens CDNA
 clone HFBBA18 similar to Alu repetitive element, mRNA sequence.
 ACCESSION M85318
 NID 9273588
 VERSION M85318.1 GI:273588
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 349)
 Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
 Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
 Sequence identification of 2,375 human brain genes
 Nature 355, 632-634 (1992)
 92168112

TITLE JOURNAL
 MEDLINE

COMMENT
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@igrr.org
 Seq primer: M13 Forward.

FEATURES
 SOURCE Location/Qualifiers
 1..349
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):77543"
 /db_xref="GDB:DO511E"
 /db_xref="taxon:9606"
 /clone="HFBBA18"
 /clone_id="Fetal brain, Striatum (cat#936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector; 1.0kb average insert size."
 96 c 73 g 101 t 7 others

BASE COUNT 72 a 96 c 73 g 101 t 7 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 349;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTATGAG 10
 |||||

DB 239 CCTATGAG 248

RESULT 8
 T01239/c

LOCUS T01239 335 bp mRNA EST 10-NOV-1992
 DEFINITION WEST01960 Early embryo, Striatum (cat. #937007) Caenorhabditis
 elegans CDNA clone CEES044, mRNA sequence.
 ACCESSION T01239
 NID 9277720
 VERSION T01239.1 GI:277720
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Pelodetidae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 335)
 McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M.,
 Fitzgerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R.,
 Venter,J.C. and Fields,C.A.
 Caenorhabditis elegans cDNAs
 Unpublished (1993)

TITLE JOURNAL
 MEDLINE

COMMENT
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@igrr.org
 Seq primer: M13 Forward.

FEATURES
 SOURCE Location/Qualifiers
 1..335
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="CEES044"
 /clone_id="Early embryo, Striatum (cat. #937007)"

BASE COUNT 116 a 67 c 64 g 85 t 3 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 335;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTATGAG 10
 |||||

DB 309 CCTATGAG 300

RESULT 9
 T03839/c

LOCUS T03839 667 bp mRNA EST 14-JUL-1993
 DEFINITION G70024 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt24
 similar to some l1 sequence similarity, mRNA sequence.
 ACCESSION T03839
 NID 9315075
 VERSION T03839.1 GI:315075
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 667)
 Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,
 Glaves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.
 A transcription map of the region containing the Huntington disease
 gene
 Hum. Mol. Genet. 2, 901-907 (1993)
 93372851

TITLE JOURNAL
 MEDLINE

COMMENT
 Contact: Hutchinson GB
 Dept. of Medical Genetics, University of British Columbia
 Canadian Genetic Diseases Network
 Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4
 Tel: (604)822-9240
 Fax: (604)822-9238
 Email: hutch@uam.genetics.ca.

FEATURES
SOURCE

Location/Qualifiers
1..667
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6 p23-p22"
/clone="gt24"
/clone.lib="Genetrack, 4p16.3 JM Rommens"

BASE COUNT
ORIGIN

187 a 120 c 98 g 262 t
/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

Query Match 100.0%; Score 10; DB 20; Length 667;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10
|||||
Db 203 CCCTATGGAG 194

RESULT 10

LOCUS T03843 422 bp mRNA EST 14-JUL-1993
DEFINITION GT0131 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt131,
MRNA sequence.

ACCESSION T03843
NID 9315082
VERSION T03843.1 GI:315082
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 422)
Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,
AUTHORS Claves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.

TITLE A transcription map of the region containing the Huntington disease
gene

JOURNAL Hum. Mol. Genet. 2, 901-907 (1993)
MEDLINE 93372851
COMMENT On Apr 14, 1993 this sequence version replaced.

Contact: Hutchinson GB
Dept. of Medical Genetics, University of British Columbia
Canadian Genetic Diseases Network
Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4
Tel: (604)822-9240
Fax: (604)822-9238
Email: hutch@hml.genes.ca.

FEATURES
SOURCE

Location/Qualifiers
1..422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="gt131"
/clone.lib="Genetrack, 4p16.3 JM Rommens"
/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

Query Match 100.0%; Score 10; DB 20; Length 667;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 157 a 108 c 107 g 50 t
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 422;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10
|||||
Db 259 CCCTATGGAG 268

RESULT 11

LOCUS T03944 236 bp mRNA EST 14-JUL-1993
DEFINITION GT1159 Genetrack, 4p16.3 JM Rommens Homo sapiens CDNA clone gt159y,
MRNA sequence.

ACCESSION T03944
NID 9315102
VERSION T03944.1 GI:315102
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 236)
Rommens,J.M., Lin,B., Hutchinson,G.B., Andrew,S.E., Goldberg,Y.P.,
AUTHORS Claves,M.L., Graham,R., Lai,V., McArthur,J., Nasir,J. et al.

TITLE A transcription map of the region containing the Huntington disease
gene
JOURNAL Hum. Mol. Genet. 2, 901-907 (1993)
MEDLINE 93372851
COMMENT

Contact: Hutchinson GB
Dept. of Medical Genetics, University of British Columbia
Canadian Genetic Diseases Network
Rm. 416, 2125 East Mall, Vancouver, B.C. V6T 1Z4
Tel: (604)822-9240
Fax: (604)822-9238
Email: hutch@hml.genes.ca.

FEATURES
SOURCE
Location/Qualifiers
1..236
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="gt159y"
/clone.lib="Genetrack, 4p16.3 JM Rommens"
/note="CDNA from frontal cortex RNA and from a pool prepared from fetal brain, frontal cortex, bone marrow and liver RNA. The cDNAs were hybridized to immobilized isolated yeast artificial chromosomes from the 4p16.3 region. Following washing at high stringency, cDNA segments that had specifically hybridized were eluted, amplified by PCR and cloned. Clones were mapped back to the YACs. Inserts were sequenced using both standard (T3 and T7) and custom primers."

BASE COUNT 39 a 48 c 54 g 88 t 7 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 236;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCCTATGGAG 10
|||||
Db 162 CCCTATGGAG 153

RESULT 12

LOCUS T04417 400 bp mRNA EST 06-NOV-1997
DEFINITION 464 Lambda-PRL2 Arabidopsis thaliana CDNA clone 40E11T7, mRNA

sequence.
 accession 104417
 NID 9315577
 VERSION 104417.1 GI:315577
 KEYWORDS EST.
 ORGANISM *thale cress*.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 1 (bases 1 to 400)
 Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H.,
 McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
 Retzel,E. and Somerville,C.
 Genes galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 Plant Physiol. 106, 1241-1255 (1994)
 95148729

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313cne@dm.cl.msu.edu
 Seq primer: T7 dye primer.
 Location/Qualifiers
 1..400
 /organism="Arabidopsis thaliana"
 /strain="var columbia"
 /db_xref="taxon:3702"
 /clone="40E117"
 /clone_1db="Lambda-PRL2"
 /note="Vector: lambda zip-lox; Site_1: Sal; Site_2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques. The vector is BR's lambda zip-lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dt primed cDNA."
 Oligo dt primed cDNA. 131 c 62 g 113 t 2 others

BASE COUNT
 ~IGIN 92 a 131 c 62 g 113 t

Query Match 100.0%; Score 10; DB 20; Length 400;
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGAG 10
 |||
 Db 315 CCTATGAG 324

RESULT 13
 LOCUS 104785 413 bp mRNA EST 06-NOV-1997
 DEFINITION 833 lambda-PRL2 Arabidopsis thaliana cDNA clone 3d17p, mRNA
 sequence.
 accession 104785
 NID 9315945
 VERSION 104785.1 GI:315945
 KEYWORDS EST.
 SOURCE *thale cress*.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313cne@dm.cl.msu.edu
 Seq primer: T7 dye primer.
 Location/Qualifiers
 1..413
 /organism="Arabidopsis thaliana"
 /strain="var columbia"
 /db_xref="taxon:3702"
 /clone="3d17p"
 /clone_1db="Lambda-PRL2"
 /note="Vector: lambda zip-lox; Site_1: Sal; Site_2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques. The vector is BR's lambda zip-lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dt primed cDNA."
 Oligo dt primed cDNA. 100 a 91 c 100 g 117 t 5 others

BASE COUNT
 ~IGIN 100 a 91 c 100 g 117 t

Query Match 100.0%; Score 10; DB 20; Length 413;
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGAG 10
 |||
 Db 79 CCTATGAG 88

RESULT 14
 LOCUS 106534 357 bp mRNA EST 30-JUN-1993
 DEFINITION EST04423 Fetal Brain, Striatum (cat#936206) Homo sapiens cDNA
 clone HFB078 similar to lymphocyte function associated antigen-3,
 mRNA sequence.
 accession 106534
 NID 9317683
 VERSION 106534.1 GI:317683
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Adams,M.D., Keriavage,A.R., Fields,C. and Venter,J.C.
 TITLE 3,400 expressed sequence tags identify diversity of transcripts
 from human brain
 JOURNAL Nature Genet. 4, 256-267 (1993)
 MEDLINE 93364420
 COMMENT Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878

Tel: 3018699056
Fax: 3018699423

Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES
Source Location/Qualifiers

1..357
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):83193"
/db_xref="taxon:9606"
/clone="HFDU78"
/note="Vector: Fetal brain, Stratagene (cat#936206)."
/note="Vector: LambdaZAP-II, 17-18 wk gestation, female;
oligo-dt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average insert size."
BASE COUNT 114 a 67 c 68 g 106 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 357;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGGAG 10
|||||
DB 341 CCTATGGAG 350

RESULT 15
T13875/c

LOCUS T13875 453 bp mRNA EST 07-JAN-1998
DEFINITION 2040 Lambda-PRL2 Arabidopsis thaliana cDNA clone 426277, mRNA
sequence.

ACCESSION T13875
NID 9430244
VERSION T13875.1 GI:430244

KEYWORDS
SOURCE

ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE
AUTHORS

1 (bases 1 to 453)
Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
Retzel,E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729

COMMENT

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313cne@lhm.cl.msu.edu

Seq primer: T7.

FEATURES
Source Location/Qualifiers

1..453
/organism="Arabidopsis thaliana"
/strain="var Columbia"
/db_xref="taxon:3702"
/clone="426277"
/clone_11b="lambda-PRL2"
/note="Vector: lambda ZAP-lox; Site_1: Sal; Site_2: Not;
lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRU's lambda ZAP-lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.
BASE COUNT 115 a 83 c 115 g 130 t 10 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 453;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTATGGAG 10
|||||
DB 205 CCTATGGAG 196

Search completed: September 17, 1999, 21:28:03
Job time: 14267 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:35 ; Search time 1962.17 Seconds
(without alignments)
16.208 Million cell updates/sec

Title: US-08-956-518a-91

Perfect score: 10

Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY_NUC

arched: 679419 segs, 1590154680 residues

Database :

GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_cm:*

4: gb_ov:*

5: gb_pal:*

6: gb_ph:*

7: gb_p11:*

8: gb_p12:*

9: gb_p13:*

10: gb_p14:*

11: gb_p15:*

12: gb_p16:*

13: gb_p17:*

14: gb_p18:*

15: gb_p19:*

16: gb_p20:*

17: gb_p21:*

18: gb_p22:*

19: gb_p23:*

20: gb_p24:*

21: gb_p25:*

22: gb_p26:*

23: gb_p27:*

24: gb_p28:*

25: gb_p29:*

26: gb_p30:*

27: gb_p31:*

28: gb_p32:*

29: gb_p33:*

30: gb_p34:*

31: gb_p35:*

32: gb_p36:*

33: gb_p37:*

34: gb_p38:*

35: gb_p39:*

36: gb_p40:*

37: gb_p41:*

38: gb_p42:*

39: gb_p43:*

40: gb_p44:*

41: gb_p45:*

42: gb_p46:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	100.0	1709	1	AAPSBADNA X64174 A. azollae p

Result No.	Score	Query Match	Length	DB ID	Description
2	100.0	100.0	1507	1	AB001778 Chlamydia
3	100.0	100.0	1507	1	AB001779 Chlamydia
4	100.0	100.0	1507	1	AB001780 Chlamydia
5	100.0	100.0	1507	1	AB001781 Chlamydia
6	100.0	100.0	1507	1	AB001782 Chlamydia
7	100.0	100.0	1507	1	AB001783 Chlamydia
8	100.0	100.0	1507	1	AB001784 Chlamydia
9	100.0	100.0	1507	1	AB001785 Chlamydia
10	100.0	100.0	1507	1	AB001786 Chlamydia
11	100.0	100.0	1507	1	AB001787 Chlamydia
12	100.0	100.0	1507	1	AB001788 Chlamydia
13	100.0	100.0	1507	1	AB001789 Chlamydia
14	100.0	100.0	1507	1	AB001790 Chlamydia
15	100.0	100.0	1507	1	AB001791 Chlamydia
16	100.0	100.0	1507	1	AB001802 Chlamydia
17	100.0	100.0	1507	1	AB001809 Chlamydia
18	100.0	100.0	1507	1	AB001810 Chlamydia
19	100.0	100.0	1507	1	AB001811 Chlamydia
20	100.0	100.0	1507	1	AB001812 Chlamydia
21	100.0	100.0	1507	1	AB001813 Chlamydia
22	100.0	100.0	1507	1	AB001814 Chlamydia
23	100.0	100.0	1507	1	AB001817 Chlamydia
24	100.0	100.0	1507	1	AB001818 Chlamydia
25	100.0	100.0	1507	1	AB001819 Chlamydia
26	100.0	100.0	1507	1	AB001820 Chlamydia
27	100.0	100.0	1507	1	AB001821 Chlamydia
28	100.0	100.0	1507	1	AB001822 Chlamydia
29	100.0	100.0	1507	1	AB001823 Chlamydia
30	100.0	100.0	1507	1	AB001824 Chlamydia
31	100.0	100.0	1507	1	AB001825 Chlamydia
32	100.0	100.0	1507	1	AB001826 Chlamydia
33	100.0	100.0	1507	1	AB001827 Chlamydia
34	100.0	100.0	1507	1	AB001828 Chlamydia
35	100.0	100.0	1507	1	AB001829 Chlamydia
36	100.0	100.0	1507	1	AB001830 Chlamydia
37	100.0	100.0	1507	1	AB001831 Chlamydia
38	100.0	100.0	1507	1	AB001832 Chlamydia
39	100.0	100.0	1507	1	AB001833 Chlamydia
40	100.0	100.0	1507	1	AB001834 Chlamydia
41	100.0	100.0	1507	1	AB001835 Chlamydia
42	100.0	100.0	1507	1	AB001836 Chlamydia
43	100.0	100.0	1507	1	AB001837 Chlamydia
44	100.0	100.0	1507	1	AB001838 Chlamydia
45	100.0	100.0	1507	1	AB001839 Chlamydia

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	100.0	1709	1	AAPSBADNA X64174 A. azollae p
2	100.0	100.0	1507	1	AB001778 Chlamydia
3	100.0	100.0	1507	1	AB001779 Chlamydia
4	100.0	100.0	1507	1	AB001780 Chlamydia
5	100.0	100.0	1507	1	AB001781 Chlamydia
6	100.0	100.0	1507	1	AB001782 Chlamydia
7	100.0	100.0	1507	1	AB001783 Chlamydia
8	100.0	100.0	1507	1	AB001784 Chlamydia
9	100.0	100.0	1507	1	AB001785 Chlamydia
10	100.0	100.0	1507	1	AB001786 Chlamydia
11	100.0	100.0	1507	1	AB001787 Chlamydia
12	100.0	100.0	1507	1	AB001788 Chlamydia
13	100.0	100.0	1507	1	AB001789 Chlamydia
14	100.0	100.0	1507	1	AB001790 Chlamydia
15	100.0	100.0	1507	1	AB001791 Chlamydia
16	100.0	100.0	1507	1	AB001802 Chlamydia
17	100.0	100.0	1507	1	AB001809 Chlamydia
18	100.0	100.0	1507	1	AB001810 Chlamydia
19	100.0	100.0	1507	1	AB001811 Chlamydia
20	100.0	100.0	1507	1	AB001812 Chlamydia
21	100.0	100.0	1507	1	AB001813 Chlamydia
22	100.0	100.0	1507	1	AB001814 Chlamydia
23	100.0	100.0	1507	1	AB001817 Chlamydia
24	100.0	100.0	1507	1	AB001818 Chlamydia
25	100.0	100.0	1507	1	AB001819 Chlamydia
26	100.0	100.0	1507	1	AB001820 Chlamydia
27	100.0	100.0	1507	1	AB001821 Chlamydia
28	100.0	100.0	1507	1	AB001822 Chlamydia
29	100.0	100.0	1507	1	AB001823 Chlamydia
30	100.0	100.0	1507	1	AB001824 Chlamydia
31	100.0	100.0	1507	1	AB001825 Chlamydia
32	100.0	100.0	1507	1	AB001826 Chlamydia
33	100.0	100.0	1507	1	AB001827 Chlamydia
34	100.0	100.0	1507	1	AB001828 Chlamydia
35	100.0	100.0	1507	1	AB001829 Chlamydia
36	100.0	100.0	1507	1	AB001830 Chlamydia
37	100.0	100.0	1507	1	AB001831 Chlamydia
38	100.0	100.0	1507	1	AB001832 Chlamydia
39	100.0	100.0	1507	1	AB001833 Chlamydia
40	100.0	100.0	1507	1	AB001834 Chlamydia
41	100.0	100.0	1507	1	AB001835 Chlamydia
42	100.0	100.0	1507	1	AB001836 Chlamydia
43	100.0	100.0	1507	1	AB001837 Chlamydia
44	100.0	100.0	1507	1	AB001838 Chlamydia
45	100.0	100.0	1507	1	AB001839 Chlamydia

source 1.1709
 /organism="Anabaena azollae"
 /strain="var. Caroliniana"
 /db_xref="taxon:1164"
 /clone_lib="EMBL3 genomic Sau3AI partial"

-35_signal 322 327
 -10_signal 345 350
 prim_transcript 358 >1495
 gene 413 1495
 /gene="psba"
 /gene="psba"
 /gene="psba"
 /note="p680-apoprotein of photosystem II"
 /codon_start=1
 /transl_table=11
 /product="D1 protein"
 /protein_id="CA45515.1"
 /db_xref="pid:938651"
 /db_xref="GI:38651"
 /db_xref="GI:38651"
 /db_xref="SWISS-PROT:P29270"
 /translation="MTTLQQRSTANVDRFCWMTSTENRIYIGMFGVLMPIPLVSA
 IACFVIAFAPVDIDGIREPAGSLYNNIGAVPSSNAIGLHFYIWEASL
 DEMLYGGYGVLFVFLIGVACVYGRMEELSLRGMRPVCVASVAATAVELI
 VPIGGSGSDGMDIGISCTENEMIVFOAEHNILMPHMGVAGFGGSLTSAMGSL
 VTSLSRETTETTESONTGKFGQEEETITVAAGYFGHLIFVASFNNRSLSHFEFLA
 AMPVIGWETALGVSTMAFNNGFNFNOSIIDSGRVIGTWADVINRANLGMVEMHER
 NNAHNPPLDAAGDVAIPVALTAPPING"

BASE COUNT 471 a 376 c 324 g 538 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1709;
 Best Local Similarity 100.0%; Pred. No. 9.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||

DB 288 TATGTTTAG 279

RESULT 2
 LOCUS AB001778 1507 bp DNA BCT 06-MAR-1998
 DEFINITION Chlamydia psittaci gene for 16S rRNA, strain 68C.
 ACCESSION AB001778
 NID 91902841
 VERSION AB001778.1 GI:1902841
 KEYWORDS 16S ribosomal RNA.
 RCE Chlamydia psittaci (strain:68C, ATCC VR-125T) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
 Tatsuomi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
 Univ., Epizootiology; 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
 2 (bases 1 to 1507)
 AUTHORS Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and
 Kikuchi, N.
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the
 16S rDNA sequence
 JOURNAL Unpublished (1997)
 FEATURES
 source 1.1507
 /organism="Chlamydia psittaci"
 /strain="68C, ATCC VR-125T"
 /db_xref="taxon:812"
 1.1507
 /product="16S rRNA"
 BASE COUNT 405 a 312 c 441 g 349 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||

DB 193 TATGTTTAG 184

RESULT 3
 LOCUS AB001779 1507 bp DNA BCT 06-MAR-1998
 DEFINITION Chlamydia psittaci gene for 16S rRNA, strain: Bud-1.
 ACCESSION AB001779
 NID 91902842
 VERSION AB001779.1 GI:1902842
 KEYWORDS 16S ribosomal RNA.
 SOURCE Chlamydia psittaci (strain: Bud-1) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
 Tatsuomi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
 Univ., Epizootiology; 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
 2 (bases 1 to 1507)
 AUTHORS Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and
 Kikuchi, N.
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the
 16S rDNA sequence
 JOURNAL Unpublished (1997)
 FEATURES
 source 1.1507
 /organism="Chlamydia psittaci"
 /strain="Bud-1"
 /db_xref="taxon:812"
 1.1507
 /product="16S rRNA"
 BASE COUNT 404 a 312 c 442 g 349 t

ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||

DB 193 TATGTTTAG 184

RESULT 4
 LOCUS AB001780 1507 bp DNA BCT 06-MAR-1998
 DEFINITION Chlamydia psittaci gene for 16S rRNA, isolate: Bud-11f.
 ACCESSION AB001780
 NID 91902843
 VERSION AB001780.1 GI:1902843
 KEYWORDS 16S ribosomal RNA.
 SOURCE Chlamydia psittaci (isolate: Bud-11f from a budgerigar) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
 Tatsuomi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
 Univ., Epizootiology; 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)


```

REFERENCE      2 (bases 1 to 1507)
AUTHORS        Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
FEATURES       location/Qualifiers
                1..1507
                /organism="Chlamydia psittaci"
                /isolate="Bud-11F from a budgerigar"
                /db_xref="taxon:812"
                rRNA
                1..1507
                /product="16S rRNA"
BASE COUNT     405 a      312 c      441 g      349 t
ORIGIN
Query Match    100.0%; Score 10; DB 1; Length 1507;
                Best Local Similarity 100.0%; Pred. No. 1e+04;
                Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
                |||||
Db      193 TATGTTTAG 184

RESULT 5
AB001781/c     1507 bp      DNA      BCT      06-MAR-1998
LOCUS          Chlamydia psittaci gene for 16S rRNA, isolate Bud-16F.
DEFINITION     AB001781
ACCESSION      91902844
VERSION        AB001781.1 GI:1902844
KEYWORDS       16S ribosomal RNA.
SOURCE         Chlamydia psittaci (isolate:Bud-16F from a budgerigar) DNA.
ORGANISM       Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 1507)
AUTHORS        Takahashi,T.
TITLE          Direct Submission
JOURNAL        Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
                Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
                Univ., Epizootiology, 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
                069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
                2 (bases 1 to 1507)
                Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
FEATURES       location/Qualifiers
                1..1507
                /organism="Chlamydia psittaci"
                /isolate="Bud-16F from a budgerigar"
                /db_xref="taxon:812"
                rRNA
                1..1507
                /product="16S rRNA"
BASE COUNT     405 a      312 c      441 g      349 t
ORIGIN
Query Match    100.0%; Score 10; DB 1; Length 1507;
                Best Local Similarity 100.0%; Pred. No. 1e+04;
                Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
                |||||
Db      193 TATGTTTAG 184

RESULT 7
AB001783/c     1507 bp      DNA      BCT      14-MAY-1999
LOCUS          Chlamydia abortus gene for 16S rRNA, strain B577.
DEFINITION     AB001783
ACCESSION      91902846
VERSION        AB001783.1 GI:1902846
KEYWORDS       16S ribosomal RNA.
SOURCE         Chlamydia abortus (strain:B577, ATCC VR-656, specific:host:Ovis
                aries) DNA.
ORGANISM       Chlamydia abortus
                Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 1507)
AUTHORS        Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
REFERENCE      2 (bases 1 to 1507)
AUTHORS        Takahashi,T.
TITLE          Direct Submission
JOURNAL        Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
                Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
                Univ., Epizootiology, 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
                069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
                Location/Qualifiers
                1..1507
                /organism="Chlamydia abortus"
                /specific_host="Ovis aries"
                /strain="B577, ATCC VR-656"
                /db_xref="taxon:83555"
                /note="submitted as Chlamydia psittaci"

```

```

DEFINITION     Chlamydia psittaci gene for 16S rRNA, isolate:Bud-5695.
ACCESSION      AB001782
NID            91902845
VERSION        AB001782.1 GI:1902845
KEYWORDS       16S ribosomal RNA.
SOURCE         Chlamydia psittaci (isolate:Bud-5695 from a budgerigar) DNA.
ORGANISM       Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 1507)
AUTHORS        Takahashi,T.
TITLE          Direct Submission
JOURNAL        Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
                Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
                Univ., Epizootiology, 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
                069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
                2 (bases 1 to 1507)
                Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
FEATURES       location/Qualifiers
                1..1507
                /organism="Chlamydia psittaci"
                /isolate="Bud-5695 from a budgerigar"
                /db_xref="taxon:812"
                rRNA
                1..1507
                /product="16S rRNA"
BASE COUNT     405 a      312 c      441 g      349 t
ORIGIN
Query Match    100.0%; Score 10; DB 1; Length 1507;
                Best Local Similarity 100.0%; Pred. No. 1e+04;
                Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
                |||||
Db      193 TATGTTTAG 184

RESULT 7
AB001783/c     1507 bp      DNA      BCT      14-MAY-1999
LOCUS          Chlamydia abortus gene for 16S rRNA, strain B577.
DEFINITION     AB001783
ACCESSION      91902846
VERSION        AB001783.1 GI:1902846
KEYWORDS       16S ribosomal RNA.
SOURCE         Chlamydia abortus (strain:B577, ATCC VR-656, specific:host:Ovis
                aries) DNA.
ORGANISM       Chlamydia abortus
                Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 1507)
AUTHORS        Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
                Kikuchi,N.
TITLE          Phylogenetic analyses of Chlamydia psittaci from birds based on the
                16S rRNA sequence
JOURNAL        Unpublished (1997)
REFERENCE      2 (bases 1 to 1507)
AUTHORS        Takahashi,T.
TITLE          Direct Submission
JOURNAL        Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
                Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
                Univ., Epizootiology, 582 Bunkyo-dai-Midorimachi, Ebetsu, Hokkaido
                069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
                Location/Qualifiers
                1..1507
                /organism="Chlamydia abortus"
                /specific_host="Ovis aries"
                /strain="B577, ATCC VR-656"
                /db_xref="taxon:83555"
                /note="submitted as Chlamydia psittaci"

```


RNA 1.1507
 /product="16S rRNA"
 BASE COUNT 408 a 312 c 438 g 349 t
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 DB 193 TATGTTTAG 184

RESULT 8
 AB001786 1507 bp DNA BCT 06-MAR-1998
 LOCUS Chlamydia psittaci gene for 16S rRNA, strain:GCP-1.
 DEFINITION
 ACCESSION
 91902849

KEYWORDS
 SOURCE Chlamydia psittaci (strain:GCP-1) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia;

REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE 2 (bases 1 to 1507)
 AUTHORS Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence
 JOURNAL Unpublished (1997)
 FEATURES Location/Qualifiers

1.1507
 /organism="Chlamydia psittaci"
 /strain="GCP-1"
 /db_xref="taxon:812"
 1.1507
 /product="16S rRNA"
 BASE COUNT 405 a 312 c 441 g 349 t
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 DB 193 TATGTTTAG 184

RESULT 9
 AB001787 1507 bp DNA BCT 06-MAR-1998
 LOCUS Chlamydia psittaci gene for 16S rRNA, strain:Itch.
 DEFINITION
 ACCESSION
 91902850
 NID
 VERSION AB001787.1 GI:1902850
 KEYWORDS 16S ribosomal RNA.
 SOURCE Chlamydia psittaci (strain:Itch) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia;
 REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission

JOURNAL

Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
 Tatsunumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen Univ., Epizootiology, 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

REFERENCE 2 (bases 1 to 1507)
 AUTHORS Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence
 JOURNAL Unpublished (1997)
 FEATURES Location/Qualifiers

1.1507
 /organism="Chlamydia psittaci"
 /strain="Itch"
 /db_xref="taxon:812"
 1.1507
 /product="16S rRNA"
 BASE COUNT 405 a 312 c 441 g 349 t
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 DB 193 TATGTTTAG 184

RESULT 10
 AB001788 1507 bp DNA BCT 06-MAR-1998
 LOCUS Chlamydia psittaci gene for 16S rRNA, isolate:Izawa-1.
 DEFINITION
 ACCESSION
 91902851
 NID
 VERSION AB001788.1 GI:1902851
 KEYWORDS 16S ribosomal RNA.
 SOURCE Chlamydia psittaci (isolate:Izawa-1 from a budgerigar) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia;

REFERENCE 1 (bases 1 to 1507)
 AUTHORS Takahashi, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE 2 (bases 1 to 1507)
 AUTHORS Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.
 TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence
 JOURNAL Unpublished (1997)
 FEATURES Location/Qualifiers

1.1507
 /organism="Chlamydia psittaci"
 /isolate="Izawa-1 from a budgerigar"
 /db_xref="taxon:812"
 1.1507
 /product="16S rRNA"
 BASE COUNT 405 a 312 c 441 g 349 t
 ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 DB 193 TATGTTTAG 184


```

RESULT 11
LOCUS AB001789/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:Koala.
ACCESSION AB001789
NID 91902852
VERSION AB001789.1 GI:1902852
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:Koala) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Koala"
/db_xref="taxon:812"
/product="16S rRNA"
BASE COUNT 405 a 312 c 441 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. NO. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
--
RESULT 12
LOCUS 01790/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, isolate:Mizuno-1F.
ACCESSION AB001790
NID 91902853
VERSION AB001790.1 GI:1902853
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (isolate:Mizuno-1F from a budgerigar) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Ohmiya"
/db_xref="taxon:812"
/product="16S rRNA"
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. NO. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
--
RESULT 13
LOCUS AB001791/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:Ohmiya.
ACCESSION AB001791
NID 91902854
VERSION AB001791.1 GI:1902854
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:Ohmiya) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="Ohmiya"
/db_xref="taxon:812"
/product="16S rRNA"
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. NO. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184
--
RESULT 14
LOCUS AB001802/c 1507 bp DNA BCT 06-MAR-1998
DEFINITION Chlamydia psittaci gene for 16S rRNA, strain:PgAu46.
ACCESSION AB001802
NID 91902865
VERSION AB001802.1 GI:1902865
KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:PgAu46) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Takahashi,T.
TITLE 1 (bases 1 to 1507)
JOURNAL Direct Submission
FEATURES
SOURCE Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.
Tatsufumi Takahashi, School of Veterinary Medicine, Rakuno Gakuen
Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido
069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)
2 (bases 1 to 1507)
Takahashi,T., Masuda,M., Tsuruno,T., Mori,Y., Takashima,I. and
Kikuchi,N.
Phylogenetic analyses of Chlamydia psittaci from birds based on the
16S rDNA sequence
Unpublished (1997)
LOCATION/Qualifiers
1..1507
/organism="Chlamydia psittaci"
/strain="PgAu46"
/db_xref="taxon:812"
/product="16S rRNA"
BASE COUNT 404 a 312 c 442 g 349 t
ORIGIN
Query Match 100.0%; Score 10; DB 1; Length 1507;
Best Local Similarity 100.0%; Pred. NO. 1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGTTTAG 10
Db 193 TATGTTTAG 184

```


TITLE Direct Submission
JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE Tatsuimi Takahashi, School of Veterinary Medicine, Rakuno Gakuen Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

AUTHORS 2 (bases 1 to 1507)
Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.

TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence
JOURNAL Unpublished (1997)

FEATURES Location/Qualifiers

source 1..1507
/organism="Chlamydia psittaci"
/strain="PgAu46"
/db_xref="taxon:812"
1..1507
/product="16S rRNA"

BASE COUNT 403 a 311 c 443 g 350 t
-GIN

Query Match 100.0%; Score 10; DB 1; Length 1507;

Best Local Similarity 100.0%; Pred. No. 1e+04; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
|||||
DB 193 TATGTTTAG 184

RESULT 15
AB001809 1507 bp DNA BCT 06-MAR-1998
LOCUS Chlamydia psittaci gene for 16S rRNA, strain:PrK46.

DEFINITION AB001809
ACCESSION 91902872
VERSION AB001809.1 GI:1902872

KEYWORDS 16S ribosomal RNA.
SOURCE Chlamydia psittaci (strain:PrK46) DNA.
ORGANISM Chlamydia psittaci

REFERENCE 1 (bases 1 to 1507)
AUTHORS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
Takahashi, T.

TITLE Direct Submission
JOURNAL Submitted (10-MAR-1997) to the DDBJ/EMBL/GenBank databases.

REFERENCE Tatsuimi Takahashi, School of Veterinary Medicine, Rakuno Gakuen Univ., Epizootiology; 582 Bunkyo-dai-Midori-machi, Ebetsu, Hokkaido 069, Japan (Tel:011-386-1112(ex.4233), Fax:011-387-5890)

AUTHORS 2 (bases 1 to 1507)
Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I. and Kikuchi, N.

TITLE Phylogenetic analyses of Chlamydia psittaci from birds based on the 16S rDNA sequence
JOURNAL Unpublished (1997)

FEATURES Location/Qualifiers

source 1..1507
/organism="Chlamydia psittaci"
/strain="PrK46"
/db_xref="taxon:812"
1..1507
/product="16S rRNA"

BASE COUNT 407 a 313 c 439 g 348 t
ORIGIN

Query Match 100.0%; Score 10; DB 1; Length 1507;

Best Local Similarity 100.0%; Pred. No. 1e+04; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
|||||
DB 193 TATGTTTAG 184

Search completed: September 17, 1999, 22:01:36
Job time: 16279 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:35 ; Search time 425.19 Seconds
(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518A-91

Perfect score: 10

Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY_NUC

arched: 311585 segs, 125096042 residues

-database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.0	9360	1	002830	CDNA to HIV-2 RNA.
2	100.0	3060	1	003098	Enod2a genomic clo
3	100.0	3856	1	003097	Enod2b genomic clo
4	100.0	18177	1	N90490	DNA of human retin
5	100.0	653	1	N90728	Total CDNA from cl
6	100.0	9822	1	N91774	Entire HIV-2/ST pr
7	100.0	774	1	N91602	Partial nucleotide
8	100.0	1975	1	N05057	Drai/Drai fragment
9	100.0	1355	1	N82025	Fragment containin
10	100.0	9643	1	N80859	Sequence of entire
11	100.0	9236	1	N80436	Entire sequence of
12	100.0	1566	1	N82198	GAGRON sequence f
13	100.0	1566	1	N70887	HIV virus gag sequ
14	100.0	4355	1	Q10947	Bovine Coronavirus
15	100.0	35	1	Q10093	Probe 782 to Chlam
16	100.0	4462	1	Q11657	3' Terminal of TSM
17	100.0	9192	1	Q11943	Nucleotide sequenc
18	100.0	40	1	Q12325	HPV type 33-specif
19	100.0	978	1	Q13847	Phibacin PSBX muta
20	100.0	983	1	Q13846	Phibacin PSBX orfl
21	100.0	1200	1	Q13848	Phibacin PSBX 1.2k
22	100.0	1201	1	Q13849	Phibacin PSBX 1.2k
23	100.0	3458	1	Q14256	51.4 and 41.9 kd t
24	100.0	2411	1	Q14658	bt1109P gene. Bac
25	100.0	15155	1	N50107	DNA encoding Facto
26	100.0	14704	1	Q20685	PKS 741 insert con
27	100.0	14704	1	Q20685	PKS 741 insert con
28	100.0	8937	1	Q20602	Net gene. DNA sequ
29	100.0	9672	1	Q20616	ROD HIV-2 isolate
30	100.0	9709	1	Q22488	HIV-1 proviral clo
31	100.0	7432	1	Q22965	True type 3 poliov
32	100.0	25	1	Q22047	Sequence of PCR pr
33	100.0	3209	1	Q22769	Canarypox PvuII fr
34	100.0	2949	1	Q25764	Canarypox virus C5
35	100.0	1105	1	Q23894	Encodes human prot
36	100.0	7323	1	Q25033	Bovine RSV strain
37	100.0	1260	1	Q22042	Sequence encoding
38	100.0	639	1	Q26962	Human PAP 3'-UTR.
39	100.0	5008	1	Q27223	MSF precursor. New
40	100.0	3209	1	Q29107	Canarypox PvuII fr
41	100.0	1521	1	Q29724	Sequence of the fi
42	100.0	7431	1	Q30148	Attenuated (4720,
43	100.0	7399	1	Q33479	HPBACR1. Defective

ALIGNMENTS

RESULT 1	002830/c	ID	002830 standard; cDNA; 9360 BP.
AC	002830	DT	29-MAY-1989 (first entry)
DE	CDNA to HIV-2 RNA.	KW	HIV, AIDS; Vaccine; pUC-HIV-2(GH-1).
OS	HIV-2.	FS	Key
FT	cds	FT	Location/Qualifiers
FT	cds	FT	544..2109
FT	cds	FT	/*tag- a
FT	cds	FT	/label-Gag reading frame
FT	cds	FT	1830..4934
FT	cds	FT	/*tag- b
FT	cds	FT	/label-Pol gene reading frame.
FT	cds	FT	4867..5511
FT	cds	FT	/*tag- c
FT	cds	FT	/label-Q gene reading frame
FT	cds	FT	5342..5677
FT	cds	FT	/*tag- d
FT	cds	FT	/label-R gene reading frame
FT	cds	FT	6148..8703
FT	cds	FT	/*tag- e
FT	cds	FT	/label-X gene reading frame
FT	cds	FT	5680..5994
FT	cds	FT	/*tag- f
FT	cds	FT	/label-Env gene reading frame.
FT	cds	FT	8540..9304
FT	cds	FT	/*tag- g
FT	cds	FT	/label-F gene reading frame.
PN	J01289486-A.	PD	21-NOV-1989.
PR	16-MAY-1988; 119024.	PA	16-MAY-1988; JP-119024.
PR	16-MAY-1988; JP-119024.	PA	Toa Nentryo Kogyo KK, Fuji Kabi KK.
DR	WPI: 90-005177/01.	DR	P-PSDB: R04024-30.
PT	DNA indicating complement to RNA gene -	PT	diagnostic for AIDS virus.
PS	Claim 2; Fig.4; 12pp; Japanese.	CC	CDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC	pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virus.	CC	Carries 7 overlapping genes in various reading frames, including gag,
CC	Pol and Env.	CC	Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T;
QY	1 TATGTTTAG 10	DB	643 TATGTTTAG 634
DB	643 TATGTTTAG 634		
RESULT 2	003098/c	ID	003098 standard; cDNA; 3060 BP.
AC	003098	DT	05-JUN-1990 (first entry)
DE	Enod2a genomic clone.	KW	Enod2; soybean; root nodule; regulatory region; nodulin 75; ss.
OS	Glycine max (L) Merr. cv. Williams.	FS	Key
FT	cds	FT	Location/Qualifiers


```

FT      exon 20      11809..11955
FT      intron (4.0 kb) 12278..12279
FT      exon 21      12538..12642
FT      intron (0.2 kb) 12948/12949
FT      exon 22      13204..13317
FT      exon 23      13411..13574
FT      intron (7.4 kb) 13883/13884
FT      exon 24      14209..14239
FT      intron (2.8 kb) 14641/14642
FT      exon 25      14834..14977
FT      exon 26      15488..15538
FT      intron (1.7 kb) 15952/15953
FT      exon 27      16143..16213
FT      polyA signal 18010

PN      M08906703-A.
PD      27-JUL-1989.
PD      23-JAN-1989; U00293.
PE      21-JAN-1988; US-146525.
      (DRYJ) Dryja T P.
      Dryja T P; Friend S; Yandell D W.
DR      WPI: 89-233856/32.
      P-PSDB: P90599.
PT      New human retino-blastoma gene and polypeptide(s)
PT      - used for screening individuals for defective retino-blastoma
PS      gene and treating these patients.
CC      Disclosure, fig 6; 71 pp. English.
CC      DNA encoding human retinoblastoma (RB) gene exons and flanking regions.
CC      The cDNA (see P90599 and N90490) can be used to screen individuals
CC      for the presence of the mutated RB gene. The RB polypeptide
CC      can prevent retinoblastoma formation, and corresp. antibodies
CC      can be used in tumour immunodiagnosis. Refer to patent for more
CC      sequence analysis details.
SQ      Sequence 18177 BP; 5765 A; 2997 C; 3172 G; 6202 T; 41 Others.

QY      1 TATGTTTAG 10
      |||||||
DB      15207 TATGTTTAG 15216

Query Match      100.0%; Score 10; DB 1; Length 18177;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
      7728
      N90728 standard; DNA; 653 BP.
      N90728;
DE      18-JUN-1990 (first entry)
DE      Total cDNA from clone for the MIP-1 beta component of murine
DE      inflammatory cytokine
DE      Inflammatory cytokine; MIP-1 beta component;
KW      mouse macrophage cell line RAW 264.7.
OS      Mouse
FH      Key      Location/Qualifiers
FT      cds      63..341
FT      mat_peptide 132..341
FT      /*tag- a
FT      /*tag- b
PN      EP-310136-A.
PD      05-APR-1989.
PD      03-OCT-1988; 116329
PR      02-SEP-1988; US-238937; US-104827.
PA      (UYRO-) Rockefeller Univ.
PI      Cerami A; Beutler B; Wolpe SD;
DR      WPI: 89-101457/14.
DR      P-PSDB: P93591.
PT      New cytokine which induces inflammation -
PT      obd. by purification of mediator substance obtd. in supernatant of
PT      stimulated macrophage cells
PS      Figure 15; 46pp; English.
CC      Cytokine MIP-1 is composed of two components, MIP-1alpha and MIP-1beta.
CC      The MIP-1 alpha was initially recovered from murine cells (mouse

```

```

CC      macrophage cell line RAW 264.7). Inflammatory cytokine may be used to
CC      identify idiopathic or invasive stimuli or determining the presence of
CC      stimulated, spontaneous or idiopathic pathological states in mammals. It
CC      induces fever in rabbits and superoxide formation in vitro in human
CC      neutrophils.
SQ      Sequence 653 BP; 139 A; 174 C; 141 G; 199 T;

Query Match      100.0%; Score 10; DB 1; Length 653;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TATGTTTAG 10
      |||||||
DB      520 TATGTTTAG 529

RESULT 6
      N91774/C
ID      N91774 standard; DNA; 9822 BP.
AC      N91774;
DE      16-MAR-1990 (first entry)
DE      Entire HIV-2/ST provirus DNA
KW      Antibodies
OS      Human immunodeficiency virus-2.
PN      W08909815-A.
PD      19-OCT-1989.
PR      04-APR-1989; U001379.
PR      04-APR-1988; US-177567.
PA      (RESF) Research Corporation Technologies Inc.
PI      Shaw GM, Hahn BH, Kong L, Lee SW;
DR      WPI: 89-324223/44.
PT      Novel virus of the HIV-2 family + used in the prodn. of diagnostic
PT      reagents, vaccines and therapies for AIDS and AIDS-related diseases
PS      Claim 37; page 47-53; 66pp; English.
CC      This is used as a source of recombinant viral components in the
CC      development of diagnostic assays for HIV-2 viruses.
SQ      Sequence 9822 BP; 3359 A; 2023 C; 2441 G; 1999 T;

QY      1 TATGTTTAG 10
      |||||||
DB      647 TATGTTTAG 638

Query Match      100.0%; Score 10; DB 1; Length 9822;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
      N91602
ID      N91602 standard; DNA; 774 BP.
AC      N91602;
DE      17-JUL-1990 (first entry)
DE      Partial nucleotide sequence (5' end) of human papilloma virus (HPV)
DE      type 33 (HPV-33)
KW      Human papilloma virus; type 33; in situ hybridisation assay;
KW      cellular smear; cervical carcinoma.
OS      Human papilloma virus.
PN      W08902934-A.
PD      06-APR-1989.
PD      30-SEP-1988; U03367.
PR      02-OCT-1987; US-103979.
PA      (MICR-) Microprobe Corp.
PI      Schwartz DE, Adams TH;
DR      WPI: 89-114406/15.
PT      Hybridisation test for human papilloma virus in cell smears -
PT      by reaction with long labelled probe specific for particular
PT      virus types; esp. for examining cervical smears
PS      Disclosure; 39pp; English.
CC      The patent is for a rapid in situ hybridisation assay for detecting and
CC      typing human papilloma virus (HPV) in non-frozen cellular smears fixed to
CC      a support in absence of aldehyde-based crosslinking reagents. The assay

```


CC comprises: (1) combining nucleic acid in the sample with at least one
 CC detectable probe able to hybridise with 1 or more HPV types; and (2)
 CC detecting presence or absence of hybrid complexes. Opt. several probes
 CC are used, eg one for HPV types 6 and 11, associated with benign warts,
 CC and one for types 16, 18, 31, 33 and 35, associated with cervical
 CC cancer. The assay can differentiate between HPV types. It is esp. used
 CC as a secondary test. The probes can be synthesised or cloned.
 SQ Sequence 774 BP; 270 A; 132 C; 163 G; 209 T.

Query Match 100.0%; Score 10; DB 1; Length 774;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||
 DB 603 TATGTTTAG 612

RESULT 8

005057 standard; DNA; 1975 BP.

DT 02-NOV-1990 (first entry)
 DE Dral/Drai fragment of the patatin gene B33 promoter.
 KW Expression cassette; transgenic potato plant; patatin gene B33;
 KW tuber-specific regulatory region; plasmid pBI 101; beta-glucuronidase;
 OS Solanum tuberosum.
 FH Key
 FH Promoter
 FT 177. 1713
 FT Location/Qualifiers
 FT misc-signal
 FT 1738. 1740
 FT /tag- a
 FT /label-Drai/Drai promoter-fragment
 FT /note- from patatin gene B33
 FT /tag- b
 FT /function-translation start

EP-375092-A.
 PN 27-JUN-1990.
 PD 18-DEC-1989; 250117.
 PR 21-DEC-1988; DE-843627.
 PA (GENB-) INST GENBIOL FORSCH.
 PI Rochoa-Sosa M, Sonnwald U, Frommer WB, Willmitzer L, Stratman M;
 DR WPI; 90-195269/26.
 PT Tuber-specific expression cassette - for prodn. of transgenic
 PT potato plants.
 PS Disclosure; p; German.
 CS 1.527 kb Dral/Drai fragment of the patatin B33 promoter (+14 to -1513)
 was inserted into plasmid pBI 101, next to the coding region of
 beta-glucuronidase and a poly-A containing nopaline-synthase terminator.
 CC The resultant expression cassette pBI 101 B33 was transferred to an
 CC agro bacterium (DBA 4404) which was used to infect potato leaves.
 SQ Sequence 1975 BP; 690 A; 265 C; 310 G; 710 T;

Query Match 100.0%; Score 10; DB 1; Length 1975;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||
 DB 334 TATGTTTAG 343

RESULT 9

ID N82025 standard; DNA; 1355 BP.

AC N82025;
 DT 22-OCT-1990 (first entry)
 DE Fragment containing rat lipocortin gene
 KW recombinant rat lipocortin; rat abdominal dropsy; ss.
 OS Rattus.
 FH Key
 FH Location/Qualifiers

FT cds 28. 1068
 FT /tag- a
 FT /product-lipocortin

J63276497-A.
 PN 14-NOV-1988.
 PD 08-MAY-1987; JP-112145.
 PR 08-MAY-1987; J6-JP-112145.
 PA (YAMA) Yamanouchi Pharm KK.
 DR WPI; 88-365616/51.
 DR P-PSDB; P82062.

PT Recombinant rat lipocortin -
 PT obtd using gene derived from cells in rat abdominal dropsy
 PS Disclosure; Japanese
 CC The coding sequence described in the features is specifically
 CC claimed. mRNA was isolated from rat abdominal dropsy cells and cDNA
 CC synthesised from it. Probes were synthesised according to the
 CC partial amino acid sequence of rat lipocortin (see N82038 and
 CC N82039). These were used to isolate plasmids contg the desired
 CC sequence of lipocortin DNA.
 CC See also P82063.
 SQ Sequence 1355 BP; 433 A; 273 C; 302 G; 347 T;

Query Match 100.0%; Score 10; DB 1; Length 1355;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 |||||
 DB 1203 TATGTTTAG 1212

RESULT 10

N80859/C

ID N80859 standard; cDNA; 9643 BP.

AC N80859;

DT 15-NOV-1990 (first entry)
 DE Sequence of entire HIV-2 ROD genome
 KW LAV-II ROD; ss.
 OS Human immunodeficiency virus ROD.

FH Key
 FH Location/Qualifiers
 FT cds 545. 2110
 FT /tag- a
 FT /note- "p80801"
 FT 1828. 4908
 FT /tag- b
 FT /note- "p81773"
 FT 4841. 5485
 FT /tag- c
 FT /note- "p81774"
 FT 5316. 5654
 FT /tag- d
 FT /note- "p81775"
 FT 5654. 5968
 FT /tag- e
 FT /note- "p81776"
 FT 5817. 6113
 FT /tag- f
 FT /note- "p81777"
 FT 6043. 6153
 FT /tag- g
 FT /note- "p81778"
 FT 6119. 8692
 FT /tag- h
 FT /note- "p81779"
 FT 8320. 8372
 FT /tag- i
 FT /note- "p81780"
 FT 8242. 8508
 FT /tag- j
 FT /note- "p81781"
 FT 8529. 9299
 FT /tag- k


```

FT      /note="P81782"
PN      WO805440-A.
PD      28-JUL-1988.
PF      15-JAN-1988; F00025.
PR      15-APR-1987; FR-005398,
PA      (INSP) Inst Pasteur(Aliz/).
PI      Allison M, Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M,
PI      Tiollais P, Chakrabarti L, Desrosiers R;
DR      WPT-88-220290/31.
DR      P-PSDB; P80801, P81773, P81774, P81775, P81776, P81777, P81778, P81779,
DR      P81780, P81781, P81782.
PT      New peptide(s) with immunological properties of HIV-2 envelope protein -
PT      having the structure of simian immune deficiency virus proteins,
PT      useful in diagnosis and of vaccine components
PS      Disclosure; Fig 1A; 86pp; French.
CC      The SQ in n80859 was deposited on 21/2/86 at the CNCM under number I-522,
CC      reference name LAV-II ROD. It is the CDNA to HIV-2 ROD genomic RNA. SQ
CC      was compared with the SQ of the genome of SIV (Mac) (n80860) to identify
CC      common regions.
        Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T;

Query Match          100.0%; Score 10; DB 1; Length 9643;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
        |||||
DB      644 TATGTTTAG 635

RESULT 11
ID      N80436/c
AC      N80436;
DT      16-DEC-1990 (first entry)
DE      Entire sequence of LAV EL I
KW      HIV; HIV-1; AIDS; diagnosis; vaccine; probe; hybridisation; ss.
OS      Lymphadenopathy associated virus EL I.
FH      Key
FT      Location/Qualifiers
FT      misc-feature
FT      1..98
FT      /tag=a
FT      /label=R
FT      99..182
FT      /tag=b
FT      /label=U5
FT      8683..9138
FT      /tag=c
FT      /label=U3
FT      9139..9236
FT      /tag=d
FT      /label=R
FT      336..1835
FT      /tag=e
FT      /label=GAG; P80884
FT      1634..4699
FT      /tag=f
FT      /label=POL; P81854
FT      4647..5222
FT      /tag=g
FT      /label=O; P81855
FT      5165..5452
FT      /tag=h
FT      /label=R; P81856
FT      5436..5651
FT      /tag=i
FT      /label=S; P81857
FT      5830..8388
FT      /tag=j
FT      /label=ENV; P81858
FT      8393..9010
FT      /tag=k
FT      /label=F; P81859

```

```

PN      WO8707906-A.
PD      30-DEC-1987.
PF      22-JUN-1987; E00326.
PR      23-JUN-1986; EP-401380.
PA      (INSP) Inst Pasteur.
PI      Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR      WPT-88-014396/02.
DR      P-PSDB; P80884, P81854, P81855, P81856, P81857, P81858, P81859.
PT      New variants of lymphadenopathy associated virus (LAV) -
PT      used for prodn. of DNA, antigens and antibodies used in
PT      diagnosis of AIDS and pre-AIDS
PS      Claim 3: Fig 7A-7J; 72pp; English.
CC      LAV EL I (n80436) and LAV MA L (n80437) were isolated from the peripheral
CC      blood lymphocytes of patients. The different AIDS virus isolates
CC      are designated by 3 letters of the patients name. Stable probes including
CC      the DNA sequences can be used for detection of the new LAV viruses or
CC      related viruses or DNA proviruses in eg biological samples. The proteins
CC      or peptides can be used for detection of antibodies induced in vivo and
CC      present in biological fluids. The DNA can also be used for the expression
CC      of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC      polypeptides can also be used for the prodn. of antibodies for the
CC      detection of proteins related to the LAV viruses, partic. for diagnosis
CC      of AIDS or pre-AIDS.
        Sequence 9236 BP; 3360 A; 1642 C; 2190 G; 2044 T;

Query Match          100.0%; Score 10; DB 1; Length 9236;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TATGTTTAG 10
        |||||
DB      435 TATGTTTAG 426

RESULT 12
ID      N82198/c
AC      N82198;
DT      05-DEC-1990 (first entry)
DE      GAGRON sequence from HIV-2 ROD corresp. to the HIV-2 gag gene.
KW      Human immunodeficiency virus-2; HIV-2; AIDS; HIV-2 ROD; gag gene;
KW      env gene; GAGRON; p16; p26; p12; ss.
OS      HIV-2 ROD.
FH      EP-269520-A.
FT      01-JUN-1988.
FT      23-NOV-1987; 402631.
FT      21-NOV-1986; US-933184.
FT      22-JAN-1987; EP-400151.
FT      PA
FT      (INSP) Inst Pasteur.
FT      Montagnier L, Charetet S, Guetard D, Alizon M, Clavel F, Guyader M,
FT      Sonigo P, Brun-Vezinet F, Rey M, Rouzoulou C, Katlama C;
FT      WPT-88-149264/22.
FT      P-PSDB; P82676.
PT      New HIV-2 retrovirus causing AIDS - and new antigenic proteins,
PT      antibodies and complementary nucleic acid sequences.
PS      Claim 5; Page 29-30; 67pp; French.
CC      HIV-2 has RNA, esp. the gag and env regions, at least 50% pref.
CC      70% homologous or even 90%, corresponding to the gag and env
CC      regions of HIV-2 ROD.
CC      GAGRON includes coding regions for p16:nucleotides 1-405;
CC      p26:nucleotides 406-1155;
CC      p12:nucleotides 1156-1566.
CC      HIV-2 retrovirus is infectious for human T4 lymphocytes and has the
CC      morphological and immunological properties of the strains CNCM T-502,
CC      -532, -642 and -643. HIV-2 causes AIDS and is distinct from HIV-1.
CC      The sequence can be used diagnostically (in hybridisation tests).
CC      See also N82199 for the ENVN sequence and N82200 for the
CC      R and U3 regions.
        Sequence 1566 BP; 536 A; 338 C; 420 G; 272 T;

Query Match          100.0%; Score 10; DB 1; Length 1566;

```


Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
DB 100 TATGTTTAG 91

RESULT 13

N70887/c
ID N70887 standard; DNA; 1566 BP.
AC N70887;
DT 28-FEB-1991 (first entry)
DE HIV virus gag sequence.
KW Immunodeficiency virus; HIV-2; AIDS; ss.
OS Homo sapiens.
PN WO8704459-A.
PD 30-JUL-1987.
PF 22-JAN-1987; 700025.
PT 22-JAN-1986; FR-000911.
PR 22-JAN-1986; FR-000910.
PR 06-FEB-1986; FR-001635.
PR 13-FEB-1986; FR-001985.
PR 03-MAR-1986; US-835228.
PR 18-MAR-1986; FR-003881.
PR 24-MAR-1986; FR-004215.
PR 06-OCT-1986; US-916080.
PR 21-NOV-1986; US-933184.
PR 22-JAN-1987; EP-400151.
PA (INSP) INST PASTEUR.
PA (MONT) MONTAGNIER L.
PI Montagnier L, Charette S, Guetard D, Clavel F, Guyader M;
PI Sonigop, Burnvezeine F;
DR P-PSDB; P70554.
PT New type of human immuno-deficiency virus, infections for T4
PT cells - and derived antigens, immunogens, monoclonal antibodies
PT and nucleic acid sequences, eg. for diagnosis of AIDS
PS Claim 5; Page 71; 116pp; French.
CC The sequence encodes the gag protein of a new HIV-2 virus genome
CC and is infectious for human T4-lymphocytes. Abs and probes may be
CC raised, useful in HIV-2 vaccination and diagnosis respectively.
CC See also N70886, N70886-89, P70550-53
CC Sequence 1566 BP; 536 A; 336 C; 420 G; 272 T;
SO

Query Match 100.0%; Score 10; DB 1; Length 1566;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
DB 100 TATGTTTAG 91

RESULT 14

Q10947
ID Q10947 standard; DNA; 4355 BP.
AC Q10947;
DT 24-MAY-1991 (first entry)
DE Bovine Coronavirus E2 gene.
KW Bovine Coronavirus; BCV; E2; E3; vaccine; ss.
OS Bovine coronavirus - Quebec isolate.
FH Key Location/Qualifiers
FT cds 14..4102
FT /*tag= a
FT /*product= E2
FT /*tag= b
FT /*note= "conserved intergenic region"
FT signal_peptide 14..64
FT /*tag= c
FT misc_rna 3929..4027

FT /*tag= d
FT /label= transmembrane-domain

FN WO9102752-A.
PD 07-MAR-1991.
PF 10-AUG-1990; CA0252.
PR 22-AUG-1989; US-397689.
PA (VETE-) VETERINARY INFECTION.
PI Parker MD, Cox GJ, Babiuk LA;
DR WPI; 91-087247/12.
DR P-PSDB; R11061.
PT Decyribonucleic acid encoding bovine coronavirus protein E2 and
PT E3 - useful as vaccine component
PS Disclosure, Fig 3 (1-6); 70pp; English.
CC E2 is one of the four genes (N, E1, E2, E3) composing BCV.
CC Immediately 5' of the E2 gene on the viral genome is the E3 gene
CC (Q10948). The E3 gene terminates 14 nucleotides upstream from
CC the E2 polypeptide initiation codon.
CC Coding sequences can either be prepared directly by synthetic
CC methods based on this sequence, or by using the sequence to design
CC oligonucleotide probes to clone coding sequences.
CC The E2 and E3 polypeptides, or fragments of these, are used as
CC subunit antigens in vaccines for protection against or ameliorating
CC BCV, without risk of infection.
SO Sequence 4355 BP; 1219 A; 702 C; 852 G; 1582 T;

Query Match 100.0%; Score 10; DB 1; Length 4355;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
DB 164 TATGTTTAG 173

RESULT 15

Q10093
ID Q10093 standard; DNA; 35 BP.
AC Q10093;
DT 13-MAR-1991 (first entry)
DE Probe 782 to Chlamydia trachomatis 16S rRNA.
KW 16S rRNA; sexually transmitted diseases; ss.
OS Synthetic.
PN WO9015159-A.
PD 13-DEC-1990.
PF 31-MAY-1990; U02989.
PR 31-MAY-1989; US-359293.
PA (GENE-) GENE-TRAK SYST.
PI Shah T, Buharin A, Williams C, Mahan D, Lane DJ, King W;
PI WPI; 91-007228/01.
PT Nucleic acid probes - specific for r-RNA or r-DNA of Chlamydia
PT trachomatis, useful in hybridisation assay for detection of e.g.
PT sexual transmitted diseases
PS Claim 3; Page 32; 35pp; English.
CC The probe is one of nine 16S rRNA-targeted probes which are specific
CC to Chlamydia trachomatis bacteria. It may be used in a variety of
CC assay systems to detect e.g. respiratory, ocular or sexually
CC transmitted diseases. See also Q10092 and Q10094-Q10110.
SO Sequence 35 BP; 12 A; 7 C; 3 G; 13 T;

Query Match 100.0%; Score 10; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
DB 7 TATGTTTAG 16

Search completed: September 18, 1999, 00:33:37
Job time: 18977 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:03; Search time 2825.05 Seconds
(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-91
Perfect score: 10
Sequence: 1 TATGTTTAG 10

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

abase :

EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	10	100.0	208	20	D20858	D20858 HUMS01839
C 2	10	100.0	191	20	D30830	D30830 HUML11042 H
C 3	10	100.0	360	20	D35619	D35619 CELK023C8F
C 4	10	100.0	259	20	D36779	D36779 CELK037A3F
C 5	10	100.0	339	20	M79084	M79084 EST01232 Su
C 6	10	100.0	403	20	M79559	M79559 WEST00096 M
C 7	10	100.0	509	20	M89400	M89400 CEL08B11 CH
C 8	10	100.0	239	20	T02576	T02576 0213C3 CBSP
C 9	10	100.0	344	20	T03104	T03104 FB24B4 Peta
C 10	10	100.0	336	20	T05062	T05062 EST02950 Fe
C 11	10	100.0	358	20	T05141	T05141 EST03029 Fe
C 12	10	100.0	385	20	T05247	T05247 EST03136 Fe
C 13	10	100.0	364	20	T09635	T09635 0211m3 gmbp
C 14	10	100.0	318	20	T10573	T10573 hpc156 Huma
C 15	10	100.0	324	20	T10574	T10574 hpc160 Huma
C 16	10	100.0	289	20	T11931	T11931 A929F Heart
C 17	10	100.0	374	20	T15644	T15644 IB1680 Infa
C 18	10	100.0	260	20	T16586	T16586 NIB1519 Nor
C 19	10	100.0	384	20	T17266	T17266 NIB608 Norm
C 20	10	100.0	285	20	T18987	T18987 b07030c Tes
C 21	10	100.0	349	20	T23480	T23480 seq3179.1-N
C 22	10	100.0	349	20	T23733	T23733 seq585 d4HB
C 23	10	100.0	292	20	T25848	T25848 EST209 BL29
C 24	10	100.0	552	20	T25873	T25873 ESTDIR93 CD
C 25	10	100.0	262	20	T28017	T28017 EST23952 Hu
C 26	10	100.0	302	20	T29131	T29131 EST70246 Hu
C 27	10	100.0	268	20	T29634	T29634 EST88290 Hu
C 28	10	100.0	316	20	T29708	T29708 EST90996 Hu
C 29	10	100.0	272	20	T30096	T30096 EST112333 H
C 30	10	100.0	288	20	T30100	T30100 EST112343 H
C 31	10	100.0	204	20	T30294	T30294 EST14376 Hu
C 32	10	100.0	325	20	T32724	T32724 EST53631 Hu
C 33	10	100.0	345	20	T36036	T36036 EST95994 Hu
C 34	10	100.0	389	20	T37196	T37196 EST102276 S
C 35	10	100.0	370	20	T37394	T37394 EST102507 S
C 36	10	100.0	313	20	T38155	T38155 EST103532 S
C 37	10	100.0	339	20	T40472	T40472 Y802B01. s2
C 38	10	100.0	384	20	T47530	T47530 Y15805. r1
C 39	10	100.0	354	20	T50974	T50974 YD71B01. s1
C 40	10	100.0	360	20	T51037	T51037 YD71B01. r1
C 41	10	100.0	390	20	T54364	T54364 Y91904. s3
C 42	10	100.0	382	20	T54798	T54798 YD42B01. s1
C 43	10	100.0	268	20	T55438	T55438 YD38B06. s1
C 44	10	100.0	182	20	Z20143	Z20143 HSRABMIG T
C 45	10	100.0	806	54	HSW009483	A1044633 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS D20858 208 bp mRNA
DEFINITION HUMS01839 Human promyelocyte Homo sapiens cDNA clone mp0405 3',
ACCESSION D20858
NID 9304678
VERSION D20858.1 GI:504678

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 208)
AUTHORS Murakawa, K., Matsubara, K., Fukushima, A., Yoshii, J. and Okubo, K.
TITLE Chromosomal assignments of 3'-directed partial cDNA sequences representing novel genes expressed in granulocytoid cells
JOURNAL Genomics 23, 379-389 (1994)
MEDLINE 95137584

REFERENCE 2 (bases 1 to 208)
AUTHORS Okubo, K., Fukushima, A., Yoshii, J., Miyama, T., Kojima, Y., Yoshinari, H., Arimoto, J. and Matsubara, K.
TITLE Gene expression of human promyelocytic cell line HL60 before and after induction of differentiation. A new application of 3'directed cDNA sequencing
JOURNAL Unpublished (1993)
COMMENT Contact: Okubo, K., Fukushima, A., Yoshii, J., Miyama, T., Kojima, Y., Yoshinari, H., Arimoto, J. and Matsubara, K.
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-Oka, Suita, Osaka 565, Japan.
Location/Qualifiers
1..208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7"
/clone="mp0405"
/note="Human promyelocyte"
/note="Female, adult, cell_line = HL60, cell_type = promyelocyte."
BASE COUNT 87 a 28 c 20 g 71 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.1e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
|||||
DB 128 TATGTTTAG 119

LOCUS 191 bp mRNA EST 08-FEB-1995
DEFINITION HML1042 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.
ACCESSION D30830
NID 9643710
VERSION D30830.1 GI:643710
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 191)
AUTHORS Sudo, K., Chinen, K. and Nakamura, Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library
JOURNAL Genomics 24, 276-279 (1995)
MEDLINE 9513017
COMMENT Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.
Location/Qualifiers

FEATURES

source 1..191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="Human fetal lung"
BASE COUNT 54 a 49 c 24 g 54 t 10 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.2e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
|||||
DB 90 TATGTTTAG 81

RESULT 3
LOCUS D35619 360 bp mRNA EST 08-AUG-1994
DEFINITION CELK02308F Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk2308 5', mRNA sequence.
ACCESSION D35619
NID 9527069
VERSION D35619.1 GI:527069
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara, Y., Mitsuk, H., Nishigaki, A., Motomachi, T., Sugimoto, A. and Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yui Kohara
Gene Library Lab
National Institute of Genetics
Yata Iiri, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbi.nig.ac.jp.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk2308"
/clone_1lb="Yui Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"
BASE COUNT 134 a 68 c 71 g 84 t 3 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.7e+04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
|||||
DB 336 TATGTTTAG 327

RESULT 4
LOCUS D36779 259 bp mRNA EST 08-AUG-1994
DEFINITION CELK037A3F Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk37a3 5', mRNA sequence.
ACCESSION D36779
NID 9528695
VERSION D36779.1 GI:528695

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE Rhabditina; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motomashi, T., Sugimoto, A. and Tabara, H.
TITLE Toward an expression map of the C. elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykoha@nig.ac.jp.

FEATURES
SOURCE Location/Qualifiers
1. 259
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK37a3"
/note="dev-stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 111 a 48 c 26 g 73 t 1 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 259;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
|||||
Db 225 TATGTTTAG 216

RESULT 5
M79084 339 bp mRNA EST 26-MAY-1992
LOCUS EST01232 Subtracted Hippocampus, Strata gene (cat. #936205) Homo
DEFINITION sapiens cDNA clone HHCPR27, mRNA sequence.
ACCESSION M79084
M79084 9273397
NID 9273397
VERSION M79084.1 GI:273397
WORDS EST.
JRCOE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 339)
AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Uterback, T.R., Nagle, D.W., Fields, C., and Venter, J.C.
TITLE Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 632-634 (1992)
MEDLINE 92168112
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.
FEATURES
SOURCE Location/Qualifiers
1. 339
/organism="Homo sapiens"
/db_xref="ANCC (inhost):78612"
/db_xref="GDB:D052053E"

/db_xref="taxon:9606"
/clone="HHCPR27"
/clone="HHCPR27"
/clone="HHCPR27"
/note="Vector: LambdaZAP-II; The hippocampus library (#4 above) was subtracted with a fibroblast cell line cDNA library (Stratagene cat. #936209; W138 lung fibroblast cell line; Oligo-dT + random primed cDNA synthesis; LambdaZAP-II vector, 1.0kb average insert size.) by the method of Sive & St. John (Nucleic Acids Res. 16:10937, 1988)."

BASE COUNT 115 a 66 c 66 g 92 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 339;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
|||||
Db 181 TATGTTTAG 190

RESULT 6
M79559 403 bp mRNA EST 30-JUN-1992
LOCUS WEST00096 Mixed stage, Strata gene (cat. #937006) Caenorhabditis
DEFINITION elegans cDNA clone CEMS03 similar to Cytochrome b, mitochondrial, mRNA sequence.
ACCESSION M79559
M79559 9271582
NID 9271582
VERSION M79559.1 GI:271582
WORDS EST.
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 403)
AUTHORS McCombie, W.R., Adams, M.D., Kelley, J.M., Fitzgerald, M.G., Uterback, T.R., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J. and Fields, C.
TITLE Caenorhabditis elegans expressed sequence tags identify gene families and potential disease gene homologues
JOURNAL Nature Genet. 1, 124-131 (1992)
MEDLINE 93250983
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.
FEATURES
SOURCE Location/Qualifiers
1. 403
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="CEMS03"
/clone="CEMS03"
/clone="CEMS03"
/note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA library. Stratagene catalog #937006. The library is oligo dT primed and directionally cloned in the Uni-ZAP XR vector."

BASE COUNT 111 a 43 c 81 g 164 t 4 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 10; DB 20; Length 403;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 DB 15 TATGTTTAG 24

RESULT 7
 M89400 509 bp mRNA EST 02-DEC-1992
 LOCUS M89400
 DEFINITION CE108B11 Chris Martin sorted cDNA library Caenorhabditis elegans
 CDNA clone cm08b11 5' similar to cytochrome b homologous peptide,
 mRNA sequence.

ACCESSION M89400
 NID 9275295
 VERSION M89400.1 GI:275295
 KEYWORDS EST
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 509)

REFERENCE
 WATERSTON R., MARTIN C., CRAXTON M., HUYNH C., COULSON A.,
 HILLIER L., DURBIN R.K., GREEN P., SHOWNKEEN R., HALLIFRAN N.,
 HAWKINS T., WILSON R., BERRS M., DU Z., THOMAS K., THIERRY-MIEG J.,
 and SULSTON J.
 A survey of expressed genes in Caenorhabditis elegans
 Nature Genet. 1, 114-123 (1992)
 93250982

TITLE JOURNAL
 MEDLINE
 COMMENT

Contact: Waterston R.H.(USA) and Sulston J.E.(UK)
 (USA) Dept. of Genetics or (UK)
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of
 Molecular Biology
 Box 8233, 4566 Scott Ave., St. Louis, MI 63110, USA, or Hills
 Road, Cambridge CB2 2QH, UK
 Tel: (USA) (314)3627072 or (UK) (0223)248011
 Fax: (USA) (314)3624137 or (UK) (0223)402008
 Email: tw@nematoide.wustl.edu or jes@mc-lmba.cambridge.ac.uk
 single read.

FEATURES
 source
 1..509
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="cm08b11"
 /clone_lib="Chris Martin sorted cDNA library"
 /lab_host="MC1061"
 /note="Vector: lambda phage SHLX2. Mixed stage
 hermaprodite cDNA library. Partially normalized by
 successively picking groups of clones that didn't
 hybridize to previously picked clones. Vector: lambda phage
 SHLX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)) Host:
 MC1061."

BASE COUNT 151 a 49 c 96 g 204 t 9 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.5e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 DB 365 TATGTTTAG 374

RESULT 8
 T02576 239 bp mRNA EST 26-MAY-1992
 LOCUS T02576
 DEFINITION 0213C3 cbspfHB3.1, Debopam Chakrabarti Plasmodium falci parum CDNA
 Clone PF0213C, mRNA sequence.

ACCESSION T02576
 NID 9271430
 VERSION T02576.1 GI:271430

KEYWORDS EST.
 SOURCE malaria parasite P. falci parum.
 ORGANISM Plasmodium falci parum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 239)

REFERENCE
 CHAKRABARTI D., REDDY G.R., DAME J.B., ALMIRA E.C., LAPLIS P.J.,
 FERL R.J., YANG T.P., ROWE T.C., and SCHUSTER S.M.
 Analysis of expressed sequence tags from Plasmodium falci parum
 Mol. Biochem. Parasitol. 66, 97-104 (1994)
 95075403

TITLE JOURNAL
 MEDLINE
 COMMENT

Contact: Dame JB
 Department of Pathobiology
 University of Florida
 2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: dame@icbr.ifas.ufl.edu
 Seq primer: 73.

FEATURES
 source
 1..239
 /organism="Plasmodium falci parum"
 /db_xref="taxon:5833"
 /clone="PF0213C"
 /clone_lib="cbspfHB3.1, Debopam Chakrabarti"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: Bluescript SK(+); Polya+ RNA, from
 asynchronous blood stage parasites of the cloned Honduran
 HB3 isolate cultured in vitro, was reverse transcribed
 using an Oligo dt-Xho I primer. Second strand was prepared
 using RNase H and DNA polymerase I. EcoR I adapters were
 ligated to the cDNA, and it was cleaved with Xho I.
 Prepared fragments were ligated into EcoR I + Xho I
 double-digested Bluescript SK(+), and transformed E. coli
 XL1-Blue.

BASE COUNT 98 a 20 c 25 g 84 t 12 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 239;
 Best Local Similarity 100.0%; Pred. No. 2e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 DB 217 TATGTTTAG 208

RESULT 9
 T03104 344 bp mRNA EST 14-FEB-1997
 LOCUS T03104/c
 DEFINITION FB24B4 Fetal brain, Stratagene Homo sapiens CDNA clone FB24B4
 3' end, mRNA sequence.

ACCESSION T03104
 NID 9314345
 VERSION T03104.1 GI:314345
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 344)

REFERENCE
 KHAN A.S., WILCOX A.S., POLYMERPOPOULOS M.H., HOPKINS J.A.,
 STEVENS T.J., ROBINSON M., ORPANA A.K. and SIKELA J.M.
 Single pass sequencing and physical and genetic mapping of human
 CDNA's
 Nature Genet. 2, 180-185 (1992)
 94258200

JOURNAL
 MEDLINE
 COMMENT

On Sep 21, 1992 this sequence version replaced gi:279243.
 Contact: Sikele JM
 Department of Pharmacology
 University of Colorado Health Sciences Center

Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
 Tel: 3032708637
 Fax: 3032707097
 Email: nikri@etally.uchsc.edu
 Seq primer: -21M13 Universal.
 Location/Qualifiers
 1..344

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):86749"
 /db_xref="GDB:D052872E"
 /db_xref="taxon:9606"
 /clone="FB24B4"
 /clone_1lb="Fetal brain, Striatogene"
 /note="Vector: Lambda ZAP II; The FB library (catalog #937201, Striatogene) was constructed by directional cloning and oligo d(T)-priming in Lambda ZAP II phage utilizing mRNA from a 17 week human fetus (total brain)."
 98 a 75 c 76 g 95 t

BASE COUNT
 ORIGIN
 Query Match
 Best Local Similarity 100.0%; Score 10; DB 20; Length 344;
 Pred. No. 1.7e+04; Mismatches 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGTTTAG 10
 1111111111
 Db 184 TATGTTTAG 175

RESULT 10
 T05062 336 bp mRNA EST 30-JUN-1993
 LOCUS EST02950 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA
 DEFINITION clone HFB055 similar to EST containing O family repeat, mRNA
 sequence.
 accession T05062
 nid 9316215
 version T05062.1 GI:316215
 keywords EST.
 source human.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 reference 1 (bases 1 to 336)
 authors Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
 title 3,400 expressed sequence tags identify diversity of transcripts
 from human brain
 journal Nature Genet. 4, 256-267 (1993)
 medline 93364420
 comment

CONTACT: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gathersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org
 Seq primer: M13-21.
 location/Qualifiers

FEATURES
 source
 1..336
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):81799"
 /db_xref="taxon:9606"
 /clone="HFB055"
 /clone_1lb="Fetal brain, Striatogene (cat#936206)"
 /note="Vector: LambdaZAP-II, 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."
 99 a 53 c 85 g 89 t 10 others

BASE COUNT
 ORIGIN
 Query Match 100.0%; Score 10; DB 20; Length 336;

Best Local Similarity 100.0%; Pred. No. 1.7e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGTTTAG 10
 1111111111
 Db 167 TATGTTTAG 176

RESULT 11
 T05141 358 bp mRNA EST 30-JUN-1993
 LOCUS EST03029 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA
 DEFINITION clone HFB021 similar to EST containing O family repeat, mRNA
 sequence.
 accession T05141
 nid 9316293
 version T05141.1 GI:316293
 keywords EST.
 source human.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 reference 1 (bases 1 to 358)
 authors Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
 title 3,400 expressed sequence tags identify diversity of transcripts
 from human brain
 journal Nature Genet. 4, 256-267 (1993)
 medline 93364420
 comment

CONTACT: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gathersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org
 Seq primer: M13-21.
 location/Qualifiers
 1..358
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):81871"
 /db_xref="taxon:9606"
 /clone="HFB021"
 /clone_1lb="Fetal brain, Striatogene (cat#936206)"
 /note="Vector: LambdaZAP-II, 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."
 111 a 51 c 104 g 88 t 4 others

Query Match 100.0%; Score 10; DB 20; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.7e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGTTTAG 10
 1111111111
 Db 255 TATGTTTAG 264

RESULT 12
 T05247/c 385 bp mRNA EST 30-JUN-1993
 LOCUS EST03136 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA
 DEFINITION clone HFB086, mRNA sequence.
 accession T05247
 nid 9316399
 version T05247.1 GI:316399
 keywords EST.
 source human.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 reference 1 (bases 1 to 385)

AUTHORS Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C.
 TITLE 3,400 expressed sequence tags identify diversity of transcripts
 JOURNAL Nature Genet. 4, 256-267 (1993)
 MEDLINE 93364420
 COMMENT

Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org

FEATURES

Source

Location/Qualifiers

1.385
 /organism="Homo sapiens"
 /db_xref="ATCC (lnhost):81973"
 /db_xref="taxon:9606"
 /clone="HEBCT86"
 /note="Vector: Fetal brain, Striatum (cat#936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dt + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."

BASE COUNT 94 a 76 c 81 g 133 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 385;
 Best Local Similarity 100.0%; Pred. No. 1.6e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 Db 133 TATGTTTAG 124

RESULT 13
 T09635 364 bp DNA EST 29-NOV-1993
 LOCUS 021lm3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic
 DEFINITION clone 021lm, mRNA sequence.
 ACCESSION T09635
 T09635
 NID 9319467
 VERSION T09635.1 GI:319467
 KEYWORDS
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 364)
 Reddy, G.R., Chakrabarti, D., Schuster, S.M., Ferl, R.J., Almitra, E.C.
 and Dame, J.B.
 Gene sequence tags from Plasmodium falciparum genomic DNA fragments
 prepared by the genease activity of mung bean nuclease
 Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)

COMMENT

Other ESTs: 021lm7
 Contact: Dame JB
 Department of Pathobiology
 University of Florida
 2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: dame@icbr.tifas.ufl.edu
 Seq primer: #3.

FEATURES

Source

Location/Qualifiers

1.364

/organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /clone="021lm"

/clone.lib="gmbpFHB3.1, G. Roman Reddy"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: pBluescript SK(+); genomic DNA, from
 asynchronous blood stage parasites of the cloned Honduran

HB3 isolate cultured in vitro, was digested with mung bean
 nuclease in the presence of 308 formamide at 50°C
 (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
 Nucleic Acids Research 16:6883-6896). The ends of the
 fragments were ligated to EcoR V-cleaved and
 dephosphorylated pBluescript SK(+). Recombinant plasmids
 transformed E. coli XL1-Blue.
 BASE COUNT 94 a 61 c 51 g 151 t 7 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 364;
 Best Local Similarity 100.0%; Pred. No. 1.7e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 Db 48 TATGTTTAG 57

RESULT 14
 T10573/c 318 bp mRNA EST 29-NOV-1993
 LOCUS hbc156 Human pancreatic islet Homo sapiens cDNA clone hbc156 5'end,
 DEFINITION mRNA sequence.
 ACCESSION T10573
 T10573
 NID G390727
 VERSION T10573.1 GI:390727
 KEYWORDS
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 318)
 Takeda, J., Yano, H., Eng, S., Zeng, Y., and Bell, G.I.
 A molecular inventory of human pancreatic islets: sequence analysis
 of 1000 cDNA clones
 Hum. Mol. Genet. 2, 1793-1798 (1993)
 94108427

COMMENT

Contact: Bell GI or Takeda J
 HHMI
 Univ. of Chicago
 5841 S. Maryland Ave., MC1028, Chicago IL 60637
 Tel: 3127029116
 Fax: 3127020271
 Email: g-bell@uchicago.edu
 Seq primer: SK primer.
 Location/Qualifiers

FEATURES
 Source 1.318

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbc156"

/note="Vector: Human pancreatic islet"
 /clone.lib="Human pancreatic islet"
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."
 BASE COUNT 114 a 66 c 58 g 80 t
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 318;
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGTTTAG 10
 |||||
 Db 154 TATGTTTAG 145

RESULT 15
 T10574/c 324 bp mRNA EST 29-NOV-1993
 LOCUS hbc160 Human pancreatic islet Homo sapiens cDNA clone hbc160 5' end,
 DEFINITION mRNA sequence.
 ACCESSION T10574
 NID 9380728
 VERSION T10574.1 GI:390728
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Euthera; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 324)
 AUTHORS Takeda,J., Yano,H., Eng,S., Zeng,Y. and Bell,G.I.
 TITLE A molecular inventory of human pancreatic islets: sequence analysis
 of 1000 cDNA clones
 JOURNAL Hum. Mol. Genet. 2, 1793-1798 (1993)
 MEDLINE 94108427
 MENT

Contact: Bell GI or Takeda J
 HHMI
 Univ. of Chicago
 5841 S. Maryland Ave., MC1028, Chicago IL 60637
 Tel: 312/7029116
 Fax: 312/7020271
 Email: g-bell@uchicago.edu
 Seq primer: SK primer.
 Location/Qualifiers

FEATURES
 source

1..324
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbc160"
 /clone_id="Human pancreatic islet"
 /note="Vector: Lambda ZAPIT; Site_1: Eco RI; Site_2: Xho
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."
 115 a 67 c 63 g 79 t

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 324;
 Best Local Similarity 100.0%; Pred. No. 1.8e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TATGTTTATG 10
 |||||||||
 158 TATGTTTATG 149

Search completed: September 17, 1999, 21:28:05
 Job time: 14269 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:37 ; Search time 425.19 Seconds

(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518a-92

Sequence: 1 CRTCACAG 10

Scoring table: IDENTITY_NUC

Database: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	100.0	7377	1 Q03005	Full length T4 CDN
2	10	100.0	6151	1 Q03006	Full length T4 CDN
3	10	100.0	660	1 N90493	A DNA sequence of
4	10	100.0	1310	1 N90334	C-terminus of the
5	10	100.0	614	1 N90502	Heavy chain antilbo
6	10	100.0	919	1 N90695	DNA sequence of po
7	10	100.0	595	1 N90696	DNA sequence of po
8	10	100.0	527	1 N91649	Reshaped light cha
9	10	100.0	150	1 N91481	Start of coding re
10	10	100.0	540	1 N91645	Heavy chain of mon
11	10	100.0	443	1 N91820	DNA sequence of th
12	10	100.0	616	1 N91463	Sequence of human
13	10	100.0	600	1 N93069	Fragment of PCR-el
14	10	100.0	5518	1 N90647	Nucleotide sequenc
15	10	100.0	5566	1 N90648	Nucleotide sequenc
16	10	100.0	5413	1 N90649	Nucleotide sequenc
17	10	100.0	5365	1 N90646	Nucleotide sequenc
18	10	100.0	608	1 N90500	Light chain antilbo
19	10	100.0	765	1 Q04040	Anti-Leu 3a heavy
20	10	100.0	4565	1 Q04077	Human 32k alveolar
21	10	100.0	2720	1 Q04081	Vector pASPCg-SV1
22	10	100.0	2761	1 Q04082	Vector pMT-Apo:GHS
23	10	100.0	4024	1 Q04027	Human angiotensin
24	10	100.0	497	1 Q04283	Modified human lym
25	10	100.0	7316	1 Q04555	Plasmid p170.2 enc
26	10	100.0	527	1 Q03290	New synthetic huma
27	10	100.0	620	1 Q04462	Total base sequenc
28	10	100.0	818	1 Q04461	M13 VPCRI with 19
29	10	100.0	4219	1 Q04786	M13 VPCRI with 19
30	10	100.0	4353	1 Q04787	Sequence encoding
31	10	100.0	972	1 Q05105	Sequence encoding
32	10	100.0	592	1 N80622	Bovine Bone Morpho
33	10	100.0	2500	1 N80631	Human Bone Morpho
34	10	100.0	3850	1 N81584	Sequence of new pl
35	10	100.0	2839	1 Q05513	1177 CDNA encoding
36	10	100.0	664	1 Q05516	Upstream sequence
37	10	100.0	646	1 Q05515	Upstream sequence
38	10	100.0	641	1 Q05514	Upstream sequence
39	10	100.0	664	1 Q05607	Plasmid p170.2 inc
40	10	100.0	7377	1 Q05608	Plasmid pBG381 inc
41	10	100.0	6151	1 Q06161	Part 1 of CDNA der
42	10	100.0	6928	1 Q06161	
43	10	100.0	6928	1 Q06161	

ALIGNMENTS

c 44 10 100.0 345 1 Q07017 HIV-1 Rev M6 mutan
45 10 100.0 6614 1 X26905 6.6 kb transcript

RESULT 1

003005
ID 003005 standard; CDNA; 7377 BP.
AC 003005; 29-MAY-1990 (first entry)
DE Full length T4 CDNA of plasmid p170-2.
KW Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.
PN M08911860-A.
PD 14-DEC-1989.
PF 08-JUN-1989; U02453.
PR 10-JUN-1988; US-204645.
PR 20-APR-1989; US-341080.
PA (GHERO-) General Hospital Corp.
PI (BIOJ) Biogen Inc.
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
DR WPI; 90-007302/01.
DR P-PSDB; R04031.
PT Combinations of soluble T4 protein and anti-retroviral agent -
PT having synergistic activity in treatment and prevention of AIDS.
PS Disclosure; h.
CC Soluble T4 constructs may be produced by truncating this sequence, to
CC remove the transmembrane and intracytoplasmic domains whilst retaining
CC the extracellular region responsible for HIV binding. The sol. T4 is
CC combined with an anti-viral agent such as AZT.
SQ See also Q03006.
Sequence 7377 BP; 1760 A; 1954 C; 1902 G; 1761 T;

Query Match 100.0%; Score 10; DB 1; Length 7377;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CRTCACAG 10
DB 962 CRTCACAG 971

RESULT 2

003006
ID 003006 standard; CDNA; 6151 BP.
AC 003006; 29-MAY-1990 (first entry)
DE Full length T4 CDNA of plasmid pBG381.
KW Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.
PN M08911860-A.
PD 14-DEC-1989.
PF 08-JUN-1989; U02453.
PR 10-JUN-1988; US-204645.
PR 20-APR-1989; US-341080.
PA (GHERO-) General Hospital Corp. (BIOJ) Biogen Inc.
PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
DR WPI; 90-007302/01.
DR P-PSDB; R04032.
PT Combinations of soluble T4 protein and anti-retroviral agent -
PT having synergistic activity in treatment and prevention of AIDS.
PS Disclosure; fig 2; 100pp; English.
CC Soluble T4 constructs may be produced by truncating this sequence, to
CC remove the transmembrane and intracytoplasmic domains whilst retaining
CC the extracellular region responsible for HIV binding. The sol. T4 is
CC combined with an anti-viral agent such as AZT.
SQ See also Q03005.
Sequence 6151 BP; 1493 A; 1589 C; 1586 G; 1483 T;

Query Match 100.0%; Score 10; DB 1; Length 6151;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||
 DB 970 CTCCTCCACAG 979

RESULT 3
 N90493
 ID N90493 standard; DNA; 660 BP.
 AC N90493;
 DT 20-OCT-1989 (first entry)
 DE A DNA sequence of a VH region gene.
 KW VH region gene; Immunoglobulin; HIV.
 OS Mus musculus
 FH Key Location/Qualifiers
 FT exon 156..201
 FT exon 278..646
 FT exon /*tag= a
 /*tag= b

PN EP-327000-A.
 PD 09-AUG-1989.
 PE 30-JAN-1989; 101583.
 PR 30-JAN-1988; JP-20255.
 PR 08-JUL-1988; JP-171385.
 PA (KAGA) The Chemo-Sero-Therapeutic Research Institute.
 PI Maeda H, Eda Y, Kimachi K, Tokiyoshi S, Matsushita S, Hattori T,
 PI Takatsuki K;
 DR P-PSDB; P90542.
 DR P-PSDB; P90542.
 PT Chimeric anti-human immune virus antibodies - contg. mouse variable
 PT regions and human constant regions for diagnosis, treatment and
 PT prevention of AIDS
 PS Disclosure; Fig 5; 33pp; English.
 CC The DNA is a VH region gene. See P90542, N90491-2, and N90495.
 SO Sequence 660 BP; 184 A; 155 C; 144 G; 177 T;

Query Match 100.0%; Score 10; DB 1; Length 660;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||
 DB 268 CTCCTCCACAG 277

JULT 4
 N90334/C
 ID N90334 standard; DNA; 1310 BP.
 AC N90334;
 DT 1-NOV-1989 (first entry)
 DE C-terminus of the superoxide dismutase C-100 fusion polypeptide
 KW Hepatitis C virus; CDNA; fusion polypeptide C100-3; superoxide dismutase;
 KW Non-A, non-B hepatitis.
 OS Pan troglodytes
 FH Key Location/Qualifiers
 FT cds 1..1166
 FT /*tag= a
 FT misc_feature 1..27
 FT /*tag= b
 FT misc_feature 28..42
 FT /*tag= c
 FT misc_feature 43..1131
 FT /*tag= d
 FT misc_feature 1132..1146
 FT /*tag= e
 FT GB212511-A.
 PN 26-JUL-1989.
 PD 18-NOV-1988; G27024.
 PR 18-NOV-1987; US-122714.

PA (CHIR-) Chiron Corporation.
 PI Houghton M, Choo QL, Kuo G;
 DR WPI; 89-215054/30.
 DR P-PSDB; P90182.
 PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
 PT of infection
 PS Disclosure; fig 36; 235pp; English.
 CC The sequence shows the sequence of the C-terminus of the superoxide
 CC dismutase (SOD)-C100 fusion polypeptide. The sequence encodes antigens
 CC which react with antibodies in patients with non-A non-B hepatitis
 CC (NANBH). The CDNA can be used to design probes, or to synthesise
 CC polypeptides, which are used to diagnose HCV-induced NANBH, to
 CC raise antibodies for immunoassay or treatment, or to produce vaccines.
 CC Misc. feature b shows the SOD C-terminal; c shows an adaptor; d shows the
 CC NANBH polypeptide; and e shows an extra sequence. See also P90182,
 CC N90303-33, and N90335-6.
 SO Sequence 1310 BP; 262 A; 388 C; 378 G; 282 T;

Query Match 100.0%; Score 10; DB 1; Length 1310;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||
 DB 1176 CTCCTCCACAG 1167

RESULT 5
 N90502
 ID N90502 standard; DNA; 614 BP.
 AC N90502;
 DT 16-NOV-1989 (first entry)
 DE Heavy chain antibody variable domain.
 KW Heavy chain variable domain; complementary determining region;
 KW antibody; Campath-1; lymphoma; immunosuppression.
 OS Rattus norvegicus
 FH Key Location/Qualifiers
 FT misc_feature 48
 FT /*tag= a
 FT misc_feature 70
 FT /*tag= b
 FT cds 110..154
 FT /*tag= c
 FT complement (110..154)
 FT /*tag= d
 FT intron 156..236
 FT /*tag= e
 FT cds 237..611
 FT /*tag= f
 FT complement (237..611)
 FT /*tag= g
 FT misc_signal 110..154
 FT /*tag= h
 FT misc_signal 237..248
 FT /*tag= i
 FT misc_feature 339..353
 FT /*tag= j
 FT misc_feature 396..452
 FT /*tag= k
 FT misc_feature 549..578
 FT /*tag= l
 FT intron 613
 FT /*tag= m
 PN EP-328404-A.
 PD 16-OCT-1989.
 PE 10-FEB-1989; 301291.
 PR 12-FEB-1988; GB-8603228.
 PA (MED) Medical Research Council.
 PI Clark MR, Riechmann L, Waldmann H, Winter GP;
 DR WPI; 89-235722/33.
 DR P-PSDB; P90917.

PT Antibody which binds to antigen Campath-1 - having complementary
 PT determining region of rat origin and foreign variable domain framework
 PS Disclousre; fig 2a; 30pp; English.
 CC The sequence encodes rat immunoglobulin heavy chain (see P90917).
 CC Misc features a and b show where RNA begins, h and l encode signal
 CC peptides, and j, k and l encode complementary determining regions 1, 2
 CC and 3 resp.
 SO Sequence 614 BP; 164 A; 156 C; 153 G; 141 T;

Query Match
 Best Local Similarity 100.0%; Score 10; DB 1; Length 614;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCACAG 10
 |||||
 Db 228 CTCCTCACAG 237

RESULT 6
 N90695
 ID N90695 standard; DNA: 919 BP.
 AC N90695;
 DT 15-JUL-1990 (first entry)
 DE DNA sequence of portion of plasmid pF8CIS containing cytomegalovirus
 DE enhancer, promoter and splice donor site and Ig variable region
 KW Plasmid pF8CIS; cytomegalovirus; enhancer; promoter;
 KW stabilising sequence; splice donor intron sequence; Ig variable region;
 KW splice acceptor sequence; human embryonic kidney cells (293); JM2.
 OS Cytomegalovirus +
 FH Key
 FT Location/Qualifiers
 FT Promoter
 FT 1..732
 FT /tag= a
 FT /note="Cytomegalovirus enhancer, promoter."
 FT 733..900
 FT /tag= b
 FT /note="stabilising sequence, ie. splice donor intron
 FT sequence, the Ig variable region intron and splice
 FT acceptor sequence."
 FT 621
 FT misc_rna
 FT 621
 FT /tag= c
 FT /note="Begin RNA"
 FT 800..900
 FT /tag= d
 FT /note="IG synthetic 100mer"
 FT 897..916
 FT /tag= e
 FT /note="PstI-ClaI converter"

FT EP-309235-A.
 PN 29-MAR-1989.
 PD 22-SEP-1988; 308784.
 PF 25-SEP-1987; US-101712.
 PR (GETH) Genentech Inc.
 PA
 PI Gorman CM;
 DR WPI; 89-095394/13.
 PT Transient expression system for recombinant proteins -
 PT encoding trans-activating host cell transfected with vector
 PS Disclousre; 44 pp; English.
 CC Expression vector pF8CIS contains the cytomegalovirus enhancer and prom-
 CC oter, the cytomegalovirus splice donor site and a portion of an intron,
 CC the Ig variable region intron and splice acceptor site, the cDNA encoding
 CC factor VIII and the SV40 polyadenylation site. It is used as a method for
 CC prodn. of a desired heterologous protein in a eukaryotic host cell.
 CC The method comprises: transfecting a eukaryotic host cell with a vector
 CC encoding trans-activating protein; transfecting the host cell with
 CC an expression vector having a sequence of a ds DNA comprising, a
 CC stabilising sequence downstream of a promoter and upstream of a DNA
 CC encoding the amino acid sequence of the desired protein, and a poly-
 CC -adenylation sequence upstream to a transcription termination site.
 CC The pref. host cells are human embryonic kidney cells (293) or JM2.
 CC The method provides useful quantities of a desired protein in a

CC relatively short period of time without having to establish continuous
 CC prodn. The transient expression system optimises the interaction between
 CC specific vector components and certain trans-activating proteins.
 SO Sequence 919 BP; 219 A; 244 C; 223 G; 233 T;

Query Match
 Best Local Similarity 100.0%; Score 10; DB 1; Length 919;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCACAG 10
 |||||
 Db 867 CTCCTCACAG 876

RESULT 7
 N90696
 ID N90696 standard; DNA: 595 BP.
 AC N90696;
 DT 15-JUL-1990 (first entry)
 DE DNA sequence of portion of plasmid pF8CIS containing SV40 enhancer and
 DE promoter, cytomegalovirus splice donor site and Ig intron
 KW Plasmid pF8CIS; cytomegalovirus; SV40; enhancer; promoter;
 KW stabilising sequence; splice donor intron sequence; Ig region;
 KW splice acceptor sequence; human embryonic kidney cells (293); JM2.
 OS Cytomegalovirus; SV40; +
 FH Key
 FT Location/Qualifiers
 FT Promoter
 FT 1..360
 FT /tag= a
 FT /note="SV40 enhancer and promoter"
 FT 361..580
 FT /tag= b
 FT /note="stabilising sequence which includes
 FT cytomegalovirus donor and intron sequence, the Ig
 FT variable region intron and splice acceptor sequence"
 FT 480..580
 FT /tag= c
 FT /note="IG synthetic 100mer"
 FT 573..590
 FT /tag= d
 FT /note="PstI-ClaI converter"

FT EP-309235-A.
 PN 29-MAR-1989.
 PD 22-SEP-1988; 308784.
 PF 25-SEP-1987; US-101712.
 PR (GETH) Genentech Inc.
 PA
 PI Gorman CM;
 DR WPI; 89-095394/13.
 PT Transient expression system for recombinant proteins -
 PT comprising trans-activating host cell transfected with vector
 PS Disclousre; 44 pp; English.
 CC Expression vector pF8CIS contains the SV40 enhancer and promoter, the
 CC cytomegalovirus splice donor site and a portion of the intron, the Ig
 CC intron and splice acceptor site, the cDNA encoding factor VIII and the
 CC SV40 polyadenylation and transcription termination sites.
 CC It is used as a method for prodn. of a
 CC desired heterologous protein in a eukaryotic host cell.
 CC The method comprises: transfecting a eukaryotic host cell with a vector
 CC encoding trans-activating protein; transfecting the host cell with
 CC an expression vector having a sequence of a ds DNA comprising, a
 CC stabilising sequence downstream of a promoter and upstream of a DNA
 CC encoding the amino acid sequence of the desired protein, and a poly-
 CC -adenylation sequence upstream to a transcription termination site.
 CC The pref. host cells are human embryonic kidney cells (293) or JM2.
 CC The method provides useful quantities of a desired protein in a
 CC prodn. The transient expression system optimises the interaction between
 CC specific vector components and certain trans-activating proteins.
 SO Sequence 595 BP; 136 A; 172 C; 146 G; 141 T;

Query Match

100.0%; Score 10; DB 1; Length 595;

Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCTCCACAG 10
DB 543 CTCTCCACAG 552

RESULT 8

N91649 standard; DNA: 527 BP.
AC N91649:
AT 10-MAR-1993 (revised)
DT 14-MAR-1990 (first entry)
DE Reshaped light chain variable domain HuVLLYS
KM Eukaryotic host.
PN EP-338745-A.
PD 25-OCT-1989.
PF 17-APR-1988; 303767.
PI 16-APR-1988; GB-009050.
PR (CELL) Celltech Ltd.
PT Owens RJ, Winter GP, Riechmann L, Yarranton GT, Bodmer MW;
DR P-PSDB; P91012.
PT Method of produ. of recombinant Fv fragments
PS - by transformed host cells contg. expression vectors
PT Disclosure; Fig. 2; 12pp; English.
SO Sequence 527 BP; 136 A; 156 C; 130 G; 105 T;

Query Match 100.0%; Score 10; DB 1; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCTCCACAG 10
DB 147 CTCTCCACAG 156

RESULT 9

N91481 standard; DNA: 150 BP.
AC N91481:
DT 28-FEB-1990 (first entry)
DE Start of coding region aequorin/NP fusion protein
KM Human immunodeficiency virus.
OS WO8908393-A.
PI 05-OCT-1989.
PR 20-MAR-1989; U01152.
PA 24-MAR-1988; US-173231.
PT (IGEN) Igen Inc.
PI Kenten JH, Casadei J, Well MJ;
DR WPI: 89-309634/42.
PT New luminiscent chimeric proteins - useful in highly sensitive
PT immunoassays, eg for HIV
PS Disclosure; page 35; 79pp; English.
CC used to express the chimeric variable 4-hydroxy-3-nitrophenyl (NP)
CC antigen/aequorin protein, in J5581 myeloma or other cells expressing
CC lambda1 light chain. This protein is used as a diagnostic.
CC Sequence 150 BP; 35 A; 38 C; 38 G; 39 T;

Query Match 100.0%; Score 10; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCTCCACAG 10
DB 118 CTCTCCACAG 127

RESULT 10

N91645 standard; DNA: 540 BP.

AC N91645:
DT 15-MAR-1990 (first entry)
DE Heavy chain of monoclonal antibody 6A4.
KM Monoclonal antibody 6A4; heavy chain; Pseudomonas aeruginosa; OMP-1.
FH Key Location/Qualifiers
FT cds 64..540
FT /*tag= a

EP-338395-A.
PD 25-OCT-1989.
PF 12-APR-1989; 106463.
PR 19-APR-1988; DE-813023.
PA (BEHM) Behringwerke.
PI Domdey H, Marger M, von Specht BU;
DR WPI: 89-310861/43.
DR P-PSDB; P93079.
PT Monoclonal antibody to Pseudomonas aeruginosa - and DNA coding for
PT variable antibody regions.
PS Claim 1; page 6; 7pp; german.
CC The sequence has a variable and constant region. Monoclonal antibody 6A4
CC reacts with the OMP-1 protein of all 19 known serotypes of P.aeruginosa.
CC It is used for therapy and diagnosis of infection, and as a carrier for
CC drugs. The antibody is IgG2a subclass.
SO Sequence 540 BP; 145 A; 144 C; 130 G; 121 T;

Query Match 100.0%; Score 10; DB 1; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCTCCACAG 10
DB 32 CTCTCCACAG 41

RESULT 11

N91820 standard; DNA: 443 BP.
AC N91820:
DT 22-MAR-1990 (first entry)
DE DNA sequence of the V and J regions of the anti-Tac heavy chain.
KM Anti-Tac monoclonal antibody; heavy chain; V region; J region;
KM Interleukin-2 receptor; human T-cells; chimeric antibody.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 38..442
FT /*tag= a
PN WO8909622-A.
PD 19-OCT-1989.
PF 13-APR-1988; U01578.
PR 15-APR-1988; US-181862.
PA (PROT-) Protein Design Labs, Inc.
PI Queen CL;
DR WPI: 89-324085/44.
DR P-PSDB; P93151.
PT New interleukin-2 receptor-specific chimeric antibodies - useful for
PT treating T-cell mediated disorders.
PS Disclosure; Fig. 2; 46pp; English.
CC The anti-Tac heavy chain is produced by ATCC CRL 9688. It is used in a
CC mouse/human chimeric antibody, which is specifically reactive with
CC interleukin-2 receptors on human T-cells.
CC Sequence 443 BP; 117 A; 114 C; 110 G; 102 T;

Query Match 100.0%; Score 10; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCTCCACAG 10
DB 6 CTCTCCACAG 15


```

RESULT 12
N91463
ID N91463 standard; DNA; 616 BP.
AC N91463;
DE 24-NOV-1989 (first entry)
DE Sequence of humanised light chain variable domain.
KM Heavy chain variable domain; complementary determining region; antibody;
KM Campath-1; lymphoma; immunosuppression.
OS Rattus norvegicus
FH Key
FH Location/Qualifiers
FT cds
FT 110..154
FT /*tag= a
FT 237..571
FT /*tag= b
PN EP-328404-A.
PD 16-OCT-1989.
PF 10-FEB-1989; 301291.
PI 12-FEB-1988; GB-8803228.
PI (MED) Medical Research Council.
PI Clark MR, Riechmann L, Waldmann H, Winter GP.
PI P-PSDB; P90917.
PI Antibody which binds to antigen Campath-1 - having complementary
PI determining region of rat origin and foreign variable domain framework
PI regions.
PS Disclosure; fig 4; 30pp; English.
CC The sequence encodes rat immunoglobulin heavy chain (see P90917).
CC Misc features a and b show where RNA begins, h and l encode signal
CC peptides, and j, k and l encode complementary determining regions 1, 2
CC and 3 resp.
SQ Sequence 616 BP; 169 A; 174 C; 142 G; 131 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 616;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 228 CTCCTCCACAG 237

RESULT 13
N93069
N93069 standard; DNA; 600 BP.
DE 29-JUN-1990 (first entry)
DE Fragment of pCT.eik.2
KM Tissue plasminogen activator; EIK variant; T7 promoter; Tcs; ds.
FH Key
FH Location/Qualifiers
FT cds
FT 444..600
FT /*tag= a
FT /label= 5' terminal of human t-PA EIK variant gene.
PN MO8810307-A.
PD 29-DEC-1988.
PF 23-JUN-1988; U02127.
PI 24-JUN-1987; US-065794.
PI (GETH) Genetech.
PI Miller HT.
PI WPI; 89-023849/03.
PI P-PSDB; P94866.
PI Recombinant protein mfr. - using balanced inducible transcriptional
PI control system, esp. for mfg. tissue plasminogen activator.
PS Disclosure; fig. 3; 48pp; English.
CC The EIK variant of t-PA is described in EP-199574 and is characterised by
CC substitutional mutations that eliminate a proteolysis site. Plasmid
CC pCT.eik.2 contains CMV enhancer/promoter region, the Ig splice acceptor,
CC the 5p promoter and RNA start, the T7 promoter, the T7E1 leader RBS, the
CC altered preprot-PA gene and the SV40 polyA site.
SQ Sequence 600 BP; 160 A; 157 C; 142 G; 141 T;

```

```

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 600;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 267 CTCCTCCACAG 276

RESULT 14
N90647
ID N90647 standard; DNA; 5518 BP.
AC N90647;
DE 01-AUG-1990 (first entry)
DE Nucleotide sequence of plasmid pBG396 including soluble T4-like (ST4)
DE polypeptide number 12 (ST4#12)
KM HIV; soluble T4-like polypeptide 12; immunotherapeutic; prophylactic;
KM plasmid pBG396; diagnostic.
OS Homo sapiens.
PN WO8901940-A.
PD 09-MAR-1989.
PI 01-SEP-1988; U02940.
PI 07-JAN-1988; US-141649, US-094322.
PI (BioJ) Bogen Inc.
PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR.
PI WPI; 89-085519/11.
PI DNA sequences coding for soluble T4-like polypeptide(s) -
PI used in immuno-therapeutic and immunosuppressive compans. and for
PI preventing, treating or detecting AIDS
PS Disclosure; 207pp; English.
CC It is the nucleotide sequence of the plasmid pBG396. The sequence
CC was isolated from 2 libraries: a lambda gt10 cDNA library derived from T
CC cell tumour line REX and a lambda gt10 cDNA library derived from
CC peripheral blood lymphocytes (PBL). For screening, a series of chemically
CC synthesised antisense oligonucleotide DNA probes based on the known T4
CC protein sequence was used. The polypeptide encoded is useful in
CC immunotherapeutic, prophylactic and diagnostic compans. It can be used
CC to purify HIV from a sample.
SQ Sequence 5518 BP; 1334 A; 1424 C; 1403 G; 1357 T;

Query Match
Best Local Similarity 100.0%; Score 10; DB 1; Length 5518;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
Db 970 CTCCTCCACAG 979

RESULT 15
N90648
ID N90648 standard; DNA; 5566 BP.
AC N90648;
DE 01-AUG-1990 (first entry)
DE Nucleotide sequence of gene encoding mature T4 sequence of plasmid pBG393
DE including soluble T4-like (ST4) polypeptide number 8 (ST4#8)
KM HIV; soluble T4-like polypeptide 8; immunotherapeutic; prophylactic;
KM plasmid pBG393; diagnostic.
OS Homo sapiens.
PN WO8901940-A.
PD 09-MAR-1989.
PI 01-SEP-1988; U02940.
PI 07-JAN-1988; US-141649, US-094322.
PI (BioJ) Bogen Inc.
PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM, Liu TR.
PI WPI; 89-085519/11.
PI DNA sequences coding for soluble T4-like polypeptide(s) -
PI used in immuno-therapeutic and immunosuppressive compans. and for
PI preventing, treating or detecting AIDS
PS Disclosure; 207pp; English.
CC It is the nucleotide sequence of region encoding the first 182 AA of the
CC mature T4 sequence of plasmid pBG393. It does not contain the additional

```


CC non-T4 6 amino acids at the C-terminus following AA 182. The sequence
CC was isolated from 2 libraries: a lambda gt10 cDNA library derived from T
CC cell tumour line REX and a lambda gt10 cDNA library derived from
CC peripheral blood lymphocytes (PBL). For screening, a series of chemically
CC synthesised antisense oligonucleotide DNA probes based on the known T4
CC protein sequence was used. The polypeptide encoded is useful in
CC immunotherapeutic, prophylactic and diagnostic compsns. It can be used
CC to purify HIV from a sample.
SQ Sequence 5566 BP; 1350 A; 1437 C; 1413 G; 1365 T; 1 Others;

Query Match

100.0%; Score 10; DB 1; Length 5566;

Best local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTCCACAG 10
|||||

DB 970 CTCTCCACAG 979

rch completed: September 18, 1999, 00:33:38
Job time: 18978 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:05 ; Search time 2825.05 Seconds

(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-92
Perfect score: 10
Sequence: 1 CNGCCACAG 10

Scoring table: IDENTITY_NDC

Search: 2546578 seqs, 98626752 residues

Database :

EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	100.0	437	5	AF119567	AF119567 Homo sapi
2	10	100.0	367	20	D15789	D15789 RICC1259A R
3	10	100.0	270	20	D21568	D21568 MUS69C02 mo
4	10	100.0	436	20	D27930	D27930 CELK0069R
5	10	100.0	253	20	D28603	D28603 MOS04G11 mo
6	10	100.0	290	20	D31189	D31189 HUM12839 H
7	10	100.0	300	20	D31275	D31275 HUM13208 H
8	10	100.0	233	20	D31359	D31359 HUM13517 H
9	10	100.0	274	20	F03188	F03188 HSC1PA122 n
10	10	100.0	239	20	L26703	L26703 MOS139A 1a
11	10	100.0	387	20	M77875	M77875 EST01459 Fe
12	10	100.0	335	20	M79545	M79545 WEST00082 M
13	10	100.0	472	20	M79810	M79810 WEST00347 M
14	10	100.0	209	20	M91496	M91496 HUMRPG6AO
15	10	100.0	353	20	T01961	T01961 WEST02682 E
16	10	100.0	400	20	T04427	T04427 474 Lambda-
17	10	100.0	466	20	T04688	T04688 735 Lambda-
18	10	100.0	174	20	T05801	T05801 EST03690 Fe
19	10	100.0	239	20	T07125	T07125 EST05014 Fe
20	10	100.0	362	20	T07194	T07194 EST05083 Fe
21	10	100.0	403	20	T08363	T08363 EST06254 In
22	10	100.0	361	20	T08639	T08639 EST06531 In
23	10	100.0	325	20	T09477	T09477 EST07370 In
24	10	100.0	399	20	T10159	T10159 seq1336 b4H
25	10	100.0	359	20	T10223	T10223 seq1333 b4H
26	10	100.0	292	20	T11222	T11222 hbc949 Huma
27	10	100.0	348	20	T11972	T11972 A843R Heart
28	10	100.0	528	20	T12189	T12189 A564F Heart
29	10	100.0	503	20	T14998	T14998 crs512 Lamb
30	10	100.0	431	20	T15177	T15177 crs661 Lamb
31	10	100.0	250	20	T16196	T16196 IB3703 Infa
32	10	100.0	370	20	T16337	T16337 NIB1145 NOR
33	10	100.0	403	20	T16939	T16939 NIB203 NOR
34	10	100.0	405	20	T18946	T18946 f11040c Tes
35	10	100.0	214	20	T23678	T23678 seq315 N-b4
36	10	100.0	259	20	T23964	T23964 seq2117 2HB
37	10	100.0	269	20	T24043	T24043 seq2201 2HB
38	10	100.0	450	20	T27084	T27084 NIBT261G03R
39	10	100.0	353	20	T27119	T27119 NIBT300E12R
40	10	100.0	461	20	T27137	T27137 NIBT314E12R
41	10	100.0	480	20	T27397	T27397 hbc2545 Hum
42	10	100.0	489	20	T28131	T28131 EST29787 Hu
43	10	100.0	379	20	T28546	T28546 EST47957 Hu
44	10	100.0	382	20	T29034	T29034 EST65948 Hu
45	10	100.0	211	54	HSM009833	A1044983 Homo sapi

ALIGNMENTS

RESULT 1
ID AF119567/c standard; RNA; EST; 437 BP.
AC AF119567;
XX
SV AF119567.1
XX

NI 94325172
 XX 11-MAR-1999 (Rel. 59, Created)
 DT 11-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX
 DE Homo sapiens clone IEDP1-2, mRNA sequence.
 XX
 KW EST.
 XX
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 CC Primates; Catarrhini; Homidae; Homo.
 XX
 XX [1]
 RN 1-437
 RP MEDLINE: 97471696.
 RX Jacob A.N., Baskaran N., Kandpal G., Narayan D., Bhargava A.K.,
 RA Kandpal R.P.;
 RT "Isolation of human ear specific cDNAs and construction of cDNA
 libraries from surgically removed small amounts of inner ear tissues";
 Somat. Cell Mol. Genet. 23(2):83-95(1997).
 -- [2]
 RN 1-437
 RP Nellisery J.K., Kandpal R.P.;
 RT Submitted (13-JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL Fels Institute for Cancer Research and Molecular Biology, Temple
 RL University School of Medicine, 3307 North Broad Street, Philadelphia, PA
 RL 19140, USA
 XX
 XX Key Location/Qualifiers
 FH 1. 437
 FT source
 FT /db_xref="taxon:9606"
 FT /note="Isolated by kinetic enrichment in a representational
 difference analysis (RDA) experiment; similar to sequence
 encoding MHC class II HLA-RP1"
 FT /organism="Homo sapiens"
 FT /tissue-type="inner and middle ear"
 FT /dev-stage="fetus"
 FT /clone="IEDP1-2"
 XX
 SQ Sequence 437 BP; 114 A; 93 C; 139 G; 77 T; 14 other;

Query Match 100.0%; Score 10; DB 5; Length 437;
 Best Local Similarity 100.0%; Pred. No. 9.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||||||||
 DB 168 CTCCTCCACAG 159

RESULT 2
 DI5789/c DI5789 367 bp mRNA EST 20-JUL-1998
 LOCUS
 DEFINITION RIC1259A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
 ACCESSION DI5789
 NID 9286983
 VERSION DI5789.1 GI:286983
 KEYWORDS EST.
 SOURCE
 ORGANISM Oryza sativa.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 367)
 Sasaki, T. and Minobe, Y.
 Rice cDNA from callus
 Unpublished (1994)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@affrc.go.jp
 PROJECT "RGP"
 Seq primer: oligo(dT).
 Location/Qualifiers
 1. 367
 /organism="Oryza sativa"
 /strain="cultivar Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone="R"
 /clone_1lb="Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus RNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid."
 BASE COUNT 53 a 114 c 127 g 69 t 4 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 367;
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||||||||
 DB 125 CTCCTCCACAG 116

RESULT 3
 DI21568 270 bp mRNA EST 07-OCT-1996
 LOCUS
 DEFINITION MOS69C02 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
 clone 69C02, mRNA sequence.
 ACCESSION DI21568
 NID 6618685
 VERSION DI21568.1 GI:618685
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 270)
 Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.
 A survey of genes expressed in undifferentiated mouse embryonal
 carcinoma F9 cells: characterization of low-abundance mRNAs
 J. Biochem. 116, 128-139 (1994)
 95096008

COMMENT
 Contact: Kazunori Shimada
 Department of Medical Genetics, Division of Molecular Biomedicine
 Research Institute for Microbial Diseases, Osaka University
 3-1, Yamadaoka, Suita, Osaka, 565, Japan
 Tel: 06-879-8325
 Fax: 06-879-8326

FEATURES
 source
 1. 270
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="69C02"
 /clone_1lb="mouse embryonal carcinoma cell line F9"
 BASE COUNT 96 a 63 c 81 g 26 t 4 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 270;
 Best Local Similarity 100.0%; Pred. No. 8.4e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||
 Db 99 CTCCTCCACAG 108

RESULT 4
 D27930/c 436 bp mRNA EST 20-NOV-1995
 LOCUS CLK006B9R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone ykcb9 3', mRNA sequence.
 ACCESSION D27930
 NID 9522671
 VERSION D27930.1 GI:522671
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
 1 (bases 1 to 436)
 Kohara, Y., Mitsuki, H., Nishigaki, A., Motoshashi, T., Sugimoto, A. and
 Tabara, H.
 Toward an expression map of the C. elegans genome
 Unpublished (1994)

TITLE JOURNAL
 COMMENT

CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers

FEATURES
 source
 1. 436
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /map="18 q21.1-q21.3"
 /clone="ykcb9"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /note="dev-stage-varied, sex-Hermaphrodite male,
 tissue-type-whole animal"

BASE COUNT 140 a 89 c 82 g 125 t

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 436;
 Best Local Similarity 100.0%; Pred. No. 9.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||
 Db 49 CTCCTCCACAG 40

RESULT 5
 D28603 253 bp mRNA EST 07-OCT-1996
 LOCUS M0504G11 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
 DEFINITION clone 04G11, mRNA sequence.
 ACCESSION D28603
 NID 9618497
 VERSION D28603.1 GI:618497
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 253)
 Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.
 A survey of genes expressed in undifferentiated mouse embryonal
 carcinoma F9 cells: characterization of low-abundance mRNAs

JOURNAL T. Biochem. 116, 128-139 (1994)
 MEDLINE 95096008
 COMMENT

CONTACT: Kazunori Shimada
 Department of Medical Genetics, Division of Molecular Biomedicine
 Research Institute for Microbial Diseases, Osaka University
 3-1, Yamadaoka, Suita, Osaka, 565, Japan
 Tel: 06-879-8325
 Fax: 06-879-8326.
 Location/Qualifiers

FEATURES
 source
 1. 253
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="04G11"
 /clone_lib="mouse embryonal carcinoma cell line F9"

BASE COUNT 56 a 82 c 58 g 50 t 7 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 253;
 Best Local Similarity 100.0%; Pred. No. 8.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||
 Db 101 CTCCTCCACAG 110

RESULT 6
 D31189 290 bp mRNA EST 08-FEB-1995
 LOCUS HUM12839 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.
 DEFINITION D31189
 ACCESSION 9644069
 NID 9644069
 VERSION D31189.1 GI:644069
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 290)
 Sudo, K., Chinen, K. and Nakamura, Y.
 2038 expressed sequence tags (ESTs) from a human fetal lung cDNA
 library
 JOURNAL Genomics 24, 276-279 (1995)
 MEDLINE 95213017
 COMMENT

CONTACT: Yusuke Nakamura
 Institute of Medical Science
 University of Tokyo
 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
 Tel: 81-3-5449-5372
 Fax: 81-3-5449-5433
 Email: yusuke@ims.u-tokyo.ac.jp
 Insert Length: 825 Std Error: 0.00
 High quality sequence stop: 1.
 Location/Qualifiers

FEATURES
 source
 1. 290
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal lung"

BASE COUNT 66 a 78 c 78 g 61 t 7 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 290;
 Best Local Similarity 100.0%; Pred. No. 8.3e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
 |||
 Db 126 CTCCTCCACAG 135


```

RESULT 7
LOCUS D31375 300 bp mRNA EST 08-FEB-1995
DEFINITION HUM13208 Human fetal lung Homo sapiens CDNA 5', mRNA sequence.
ACCESSION D31375
NID 9644155
VERSION D31275.1 GI:644155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 300)
AUTHORS Sudo,K., Chinen,K. and Nakamura,Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
JOURNAL GENOMICS 24, 276-279 (1995)
MEDLINE 95213017
ENT

Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp
Insert Length: 2137 Std Error: 0.00
High quality sequence stop: 377.
Location/Qualifiers
1. 300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="710B10; 15"
/clone_lib="Human fetal lung"

BASE COUNT 70 a 83 c 82 g 63 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.6e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
Db 128 CTCCTCCACAG 137

...SULT 8
LOCUS D31359 233 bp mRNA EST 08-FEB-1995
DEFINITION HUM13117 Human fetal lung Homo sapiens CDNA 5', mRNA sequence.
ACCESSION D31359
NID 9644239
VERSION D31359.1 GI:644239
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 233)
AUTHORS Sudo,K., Chinen,K. and Nakamura,Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
JOURNAL GENOMICS 24, 276-279 (1995)
MEDLINE 95213017
ENT

Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433

```

```

Email: yusuke@ims.u-tokyo.ac.jp
Insert Length: 586 Std Error: 0.00
High quality sequence stop: 400.
Location/Qualifiers
1. 233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal lung"

BASE COUNT 45 a 72 c 59 g 55 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 233;
Best Local Similarity 100.0%; Pred. No. 8.1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
Db 192 CTCCTCCACAG 201

RESULT 9
LOCUS F03188/c 274 bp mRNA EST 02-FEB-1995
DEFINITION HSC1PA122 normalized infant brain cDNA Homo sapiens CDNA clone
c-1pa12.3', mRNA sequence.
ACCESSION F03188
NID 9646745
VERSION F03188.1 GI:646745
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 274)
AUTHORS Devignes,M.D., Duprat,S., Houllatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Sanson,R., Pietu,G., Pouillot,Y.,
Sebastiani-Kabackich,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534

JOURNAL GENOMICS 24, 276-279 (1995)
MEDLINE 95277534
COMMENT

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. removed at sequence 5'end
Genexpress_library_idt: C; Genexpress_sequence_idt: alc-1pa12
Seq primer: (-21)M13 universal.
Location/Qualifiers
1. 274
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="X q24-qter"
/clone="c-1pa12"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site:1: HindIII;
Site:2: NotI; sex:Female; dev:stage=3 months old;
isolate-muscular atrophy patient; tissue-type=total
brain; total mRNA was oligo (dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 65 a 70 c 53 g 86 t

```


ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 274;
 Best Local Similarity 100.0%; Pred. No. 8.4e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||

Db 180 CTCCTCCACAG 171

RESULT 10
 LOCUS L26703 239 bp mRNA EST 01-MAR-1995
 DEFINITION M05F139A lambda unizap male testis Mus musculus cDNA clone F139,
 mRNA sequence.
 ACCESSION L26703
 "ID 9437492
 L26703.1 GI:437492
 EST.
 "XREFS house mouse;
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 239)
 AUTHORS Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 TITLE Analysis of cDNA sequences from mouse testis
 JOURNAL Mamm. Genome 5, 557-565 (1994)
 MEDLINE 95093181
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 Location/Qualifiers
 1..239
 /organism="Mus musculus"
 /strain="Swiss"
 /db_xref="taxon:10090"
 /clone_lib="F139"
 /note="Vector: lambda unizap; lambda unizap library from
 mail mouse testis"

BASE COUNT 66 a 62 c 78 g 32 t 1 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 239;
 Best Local Similarity 100.0%; Pred. No. 8.2e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||

Db 180 CTCCTCCACAG 171

RESULT 11
 LOCUS M77875 387 bp mRNA EST 26-MAY-1992
 DEFINITION EST01459 Fetal brain, strata gene (cat#936206) Homo sapiens cDNA
 clone HFBCA49 similar to Ribosomal protein Y10, mRNA sequence.
 ACCESSION M77875
 "ID 9273613
 M77875.1 GI:273613
 EST.
 "XREFS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 387)
 AUTHORS Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
 Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112

COMMENT

Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..387
 /organism="Homo sapiens"
 /db_xref="ATCC (host):77646"
 /db_xref="GDB:D05885E"
 /db_xref="taxon:9606"
 /clone_lib="HFBCA49"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dt + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0Kb average insert size."

BASE COUNT 87 a 100 c 104 g 94 t 2 others

ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 387;
 Best Local Similarity 100.0%; Pred. No. 9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCCACAG 10
 |||||

Db 96 CTCCTCCACAG 105

RESULT 12
 LOCUS M79545 335 bp mRNA EST 30-JUN-1992
 DEFINITION W05700082 Mixed stage, Strata gene (cat. #937006) Caenorhabditis
 elegans cDNA clone CEMS452 similar to 3-Methyl-2-oxobutanoate
 dehydrogenase homologous peptide, mRNA sequence.
 ACCESSION M79545
 "ID 9271570
 M79545.1 GI:271570
 EST.
 "XREFS Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 335)
 AUTHORS McCombie,M.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G.,
 Utterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and
 Fields,C.
 TITLE Caenorhabditis elegans expressed sequence tags identify gene
 families and potential disease gene homologues
 JOURNAL Nature Genet. 1, 124-131 (1992)
 MEDLINE 93250983
 COMMENT Other ESTs: W05700081
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..335
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone_lib="CEMS452"
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA
 library. Strata gene catalog #937006. The library is oligo
 dt primed and directionally cloned in the Uni-ZAP XR

BASE COUNT 96 a 64 c 88 g 80 t 7 others
ORIGIN vector."

Query Match 100.0%; Score 10; DB 20; Length 335;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
|||||
DB 225 CTCCTCCACAG 216

RESULT 13
M79810/c 472 bp mRNA EST 30-JUN-1992
LOCUS W5100347 Mixed stage, StrataGene (cat. #937006) Caenorhabditis
DEFINITION elegans CDNA clone CEMSE18, mRNA sequence.
VERSION M79810.1 GI:271829
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditioidae; Rhabditiidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 472)
AUTHORS McCombie, W.R., Adams, M.D., Kelley, J.M., Fitzgerald, M.G.,
Uterback, T.R., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J. and
Fields, C.
Caenorhabditis elegans expressed sequence tags identify gene
families and potential disease gene homologues
Nature Genet. 1, 124-131 (1992)
93250983

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@ligr.org
Seq primer: M13 Forward.
Location/Qualifiers
1. 472
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="Mixed stage, StrataGene (cat. #937006)"
/note="Vector: Uni-ZAP XR. C. elegans mixed stage cDNA
library. StrataGene catalog #937006. The library is oligo
dT primed and directionally cloned in the Uni-ZAP XR
vector."

BASE COUNT 123 a 112 c 95 g 141 t 1 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 472;
Best Local Similarity 100.0%; Pred. No. 9.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
|||||
DB 444 CTCCTCCACAG 435

RESULT 14
M91496 209 bp mRNA EST 29-OCT-1992
LOCUS HUMR1PGE80 Subtracted human retinal pigment epithelium (RPE) Homo
DEFINITION sapiens CDNA, mRNA sequence.
ACCESSION M91496

NID 9275912
VERSION M91496.1 GI:275912
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 209)
AUTHORS Gieser, L. and Swaroop, A.
Expressed sequence tags and chromosomal localization of cDNA clones
from a subtracted retinal pigment epithelium library
Genomics 13, 873-876 (1992)
92347897

COMMENT

Contact: Swaroop, A.
Department of Ophthalmology
Kellogg Eye Center, University of Michigan
1000 Wall Street, Ann Arbor, MI 48105.
Location/Qualifiers
1. 209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Subtracted human retinal pigment epithelium
(RPE)"

FEATURES

source

BASE COUNT 71 a 48 c 38 g 50 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 209;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCCACAG 10
|||||
DB 97 CTCCTCCACAG 106

RESULT 15
T01961 353 bp mRNA EST 10-NOV-1992
LOCUS W5102682 Early embryo, StrataGene (cat. #937007) Caenorhabditis
DEFINITION elegans CDNA clone CEMSx81, mRNA sequence.
ACCESSION T01961
MID 9278442
VERSION T01961.1 GI:278442
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditioidae; Rhabditiidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 353)
AUTHORS McCombie, W.R., Kelley, J.M., Aubin, L., Goscochea, M.,
Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R.,
Venter, J.C. and Fields, C.A.
Caenorhabditis elegans CDNAs
Unpublished (1993)

TITLE
JOURNAL
COMMENT

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@ligr.org
Seq primer: M13 Forward.
Location/Qualifiers
1. 353
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="CEMSx81"
/clone_lib="Early embryo, StrataGene (cat. #937007)"

BASE COUNT 124 a 38 c 87 g 99 t 5 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 353;
Best Local Similarity 100.0%; Pred. No. 8.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 1 CTCCTCCACAG 10
|||||||
Db 279 CTCCTCCACAG 270

Search completed: September 17, 1999, 21:28:06
Job time: 14270 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:33:38 ; Search time 425.19 Seconds
(without alignments)
5.884 Million cell updates/sec

Title: US-08-956-518a-93

Perfect score: 10

Sequence: 1 GTCTCCCCAG 10

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	4885	003618	Phosphoenolpyruvat
2	10	100.0	1845	N90601	HindIII fragment o
3	10	100.0	829	N90234	Gastrin recombinan
4	10	100.0	2045	N90749	Human colony stimu
5	10	100.0	5406	N93197	Fragment of clone
6	10	100.0	2024	003143	Nucleotide sequenc
7	10	100.0	2743	004799	Recombinant sequen
8	10	100.0	1071	N81654	LVD regions of hu
9	10	100.0	1399	006280	Sequence encoding
10	10	100.0	15464	011415	Ryanodine receptor
11	10	100.0	1196	N91627	Sequence of gamma-
12	10	100.0	14928	011707	Immunoglobulin D-r
13	10	100.0	861	N60588	First half of part
14	10	100.0	549	N60295	DNA encoding hepat
15	10	100.0	4382	012759	P40 genomic DNA. N
16	10	100.0	973	013171	Retinal cGMP phosph
17	10	100.0	2156	013241	Human HSF cDNA seq
18	10	100.0	12364	013547	ACV5 gene pcDNA. M
19	10	100.0	12364	013607	ACV synthetase gen
20	10	100.0	3014	014954	Human dopamine DI
21	10	100.0	15155	N50107	DNA encoding Facto
22	10	100.0	1516	N50360	Endoglucanase ENDO
23	10	100.0	1696	N50179	Sequence of endogl
24	10	100.0	1170	028217	Sequence of tuft3 g
25	10	100.0	4258	022439	DNA of hgeco-1, en
26	10	100.0	5136	023313	DNA encoding mask1
27	10	100.0	13875	020240	Human centromeric
28	10	100.0	552	N00001	Sequence encoding
29	10	100.0	2743	N00003	Sequence of a part
30	10	100.0	2156	025713	Sequence of human
31	10	100.0	1191	027829	Fetal calf TP-like
32	10	100.0	1282	029958	5-Substid. hydrantio
33	10	100.0	1317	031878	Cyclin D1 promoter
34	10	100.0	3980	036780	Human cardiac cgr
35	10	100.0	2560	035298	Murine TC-CSF. New
36	10	100.0	973	043549	Gamma subunit of h
37	10	100.0	13747	047957	Myotonic dystrophy
38	10	100.0	4099	047957	Human endothelial
39	10	100.0	397	060714	Human brain expres
40	10	100.0	338	060300	Human vitamin D re
41	10	100.0	1399	051424	Human cyclin D1 pr
42	10	100.0	1316	053210	Neurocan DNA. Euka
43	10	100.0	5191	057710	

ALIGNMENTS

44 10 100.0 1593 1 053404
45 10 100.0 116277 1 X20249

Eleusine indica L
Borrelia burgdorferi

RESULT 1
ID 003618/c
AC 003618:
DT 01-AUG-1990 (first entry)
DE Phosphoenolpyruvate carboxylase (ppc) gene inserted into plasmid pTG1200.
KW Phosphoenolpyruvate carboxylase; ppc; pTG1200; ds.
OS Corynebacterium glutamicum.
FH Key location/Qualifiers
FT cds 921..3677
FT /*tag= a
PD EP-358940-A.
PD 21-MAR-1990.
PF 8-AUG-1989; 114632.
PR 12-SEP-1988; GB-021319.
PA (DECS) Degussa AG.
PI Bachmann B, Thierbach GD, Kalinowski J, Puhler A, O'Reagan M,
PI Viret JF, Lepage P, Lemoine Y;
DR P-PDB: R05512.
PT Deoxyribonucleic acid fragment, for L-amino acid -
PT obtd. from corynebacterium glutamicum strain coding for
PT phosphoenolpyruvate carboxylase.
PS Disclosure; P: English.
CC PPC is important as a key enzyme in the anaerobic function of the cell,
CC by increasing oxaloacetate levels. It indirectly increases biosynthetic
CC levels of L-IAs.
SQ Sequence 4885 BP; 1071 A; 1356 C; 1318 G; 1139 T; 1 Others;

Query Match 100.0%; Score 10; DB 1; Length 4885;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTCCCCAG 10
DB 4295 GTCTCCCCAG 4286

RESULT 2
ID N90601
AC N90601:
DT 05-JAN-1990 (first entry)
DE HindIII fragment of human CSF-1.
KW Colony stimulating factor; Immunosuppression; lymphokine.
OS Homo sapiens
FH Key location/Qualifiers
FT exon 218..340
FT /*tag= a
FT /*tag= b
FT /*tag= .1449
FT exon 1387..1449
FT /*tag= b
FT /*tag= .822
FT misc-feature /*tag= c
PD US4847201-A.
PD 11-SEP-1989.
PR 09-FEB-1988; 157094.
PR 05-FEB-1985; US-876819.
PA (Cetus) Cetus Corp.
PI Kawasaki ES, Lachner MB, Van Arsdel JN, Wang AM, Ralph P, Coyen MY,
DR WPI: 89-255732/35.
DR P-PDB: P01458.
PT Human and murine colony stimulating factors, CSF-1 - used to overcome
PT immunosuppression caused by eg chemotherapy, and obtd. by recombinant
PT methods.
PS Disclosure; fig 4a-4c; 30pp; English.

CC Feature c represents 200 unnamed bases. Source of mRNA was human
 CC pancreatic carcinoma cell line MIPaCa-2. See also P91458 and P91459.
 SO Sequence 1845 BP; 410 A; 456 C; 513 G; 466 T;

Query Match 100.0%; Score 10; DB 1; Length 1845;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 Db 525 GTCCTCCAG 534

RESULT 3
 ID N90234 standard; DNA; 829 BP.
 AC N90234;
 DT 1-NOV-1989 (first entry)
 Gastrin recombinant DNA.
 Recombinant DNA: promoter; signal peptide; gastrin; shuttle
 vector; yeast; Escherichia coli.

Key Location/Qualifiers
 exon 1..472
 intron 473..602
 exon 603..829
 intron 829..881
 exon 881..921
 intron 921..935
 exon 935..935

FT misc_feature 288..291
 FT misc_feature 320..325
 FT misc_feature 436..441
 FT misc_feature 711..716
 FT misc_feature 768..771
 FT misc_feature 794..797
 FT misc_feature 262..694
 FT cds 262..694
 FT /*tag- k
 PN J64002579-A.
 03-MAY-1989.
 24-JUN-1987; 156745.
 24-JUN-1987; JP-156745.
 PA (AJIN) Ajinomoto KK.
 DR WPI; 89-176418/24.
 DR P-PSDB; P90331.
 PT Recombinant DNA encoding gastrin - also comprising promoter and signal
 peptide coding regions. Inserted into shuttle vector to produce
 PT gastrin in yeast or E. coli.
 PS Claim: page 524 and fig 2: 10pp. Japanese.
 CC The recombinant DNA effectively produces gastrin in high quantities.
 CC It is made by integrating DNA fragment encoding gastrin into a shuttle
 CC vector which can grow in yeast or E. coli, eg pMW82, CY37, PA177, YEP51,
 CC YEP52 or pJDB218. The DNA encoding signal peptide is, eg from human
 CC gastrin, Mfalphei (yeast secretion hormone), yeast invertase, killer
 CC toxin, yeast alkaline or acid phosphatase, human lysozyme, human alpha-
 CC amylase or human trypsin inhibitor. DNA encoding gastrin is from, eg
 CC pMW53. See also P90331. Misc. feature d is Sau3A site, e is Sau3A site,
 CC f is HindIII site, g is PvuII site, h is HindIII site, i is Sau3A site,
 CC and j is Sau3A site. The CDS is minus the intron.
 CC Bases 262-318 encode signal peptide, and bases 319-429 encode gastrin.
 SO Sequence 829 BP; 199 A; 251 C; 212 G; 167 T;

Query Match 100.0%; Score 10; DB 1; Length 829;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 Db 524 GTCCTCCAG 533

RESULT 4
 ID N90749 standard; CDNA; 2045 BP.
 AC N90749;
 DT 06-JUN-1990 (first entry)
 DE Human colony stimulating factor-1 (CSF-1)
 DE Human colony stimulating factor-1; CSF-1; cytomagalovirus; CMV;
 KW Gram-negative sepsis; sarcoma therapy; melanoma therapy; ADCC;
 KW antibody-dependent cellular cytotoxicity; wound healing.
 OS Homo sapiens.
 FH Key
 FT Introns
 FT 1..217
 FT /*tag- a
 FT 218..340
 FT /*tag- b
 FT 341..1581
 FT /*tag- c
 FT 1582..1650
 FT /*tag- d
 PN W08902746-A.
 PD 06-APR-1989.
 PF 16-SEP-1988; 003234.
 PR 14-SEP-1988; US-243253, US-099872.
 PA (CERV) Cetus Corp.
 PI Ralph P. Warren MK, Chong KT, Devlin JT, Zimmerman R;
 DR WPI; 89-114251/15.
 DR P-PSDB; P93555; P93556.
 PT Use of recombinant colony stimulating factor-1 -
 PT for enhancing prodn. of white blood cells or treating viral or
 PT bacterial infectious diseases or tumours
 PS Figure 2a-2c; pp. 3/7-5/7; 62pp; English.
 CC It is the sequenced portion of a 3.9kb HindIII fragment encoding human
 CC CSF-1 sequences. It is used in the mfr. of a compsn. useful for
 CC enhancing prodn. of white blood cells, including neutrophils, from stem
 CC cells. It is also used in mfr. of a compsn. useful for treatment/
 CC prophylaxis of bacterial or viral infectious diseases (esp.
 CC cytomegalovirus or Gram-negative sepsis), for the treatment
 CC of sarcoma or melanoma, for stimulating antibody-dependent cellular
 CC cytotoxicity (ADCC), and for wound healing.
 SO Sequence 2045 BP; 410 A; 455 C; 514 G; 466 T; 200 Others;

Query Match 100.0%; Score 10; DB 1; Length 2045;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 Db 525 GTCCTCCAG 534

RESULT 5
 ID N93197 standard; DNA; 5406 BP.
 AC N93197;
 DT 28-JUN-1990 (first entry)
 DE Fragment of clone lambda HK65a contg. kallikrein gene.
 DE Kallikrein gene; vasodilator; male infertility; lambda HK65a; ss.
 KW Homo sapiens.
 OS Homo sapiens.
 FH Key
 FT exon 803..848
 FT /*tag- a
 FT 2538..2697
 FT /*tag- b
 FT 3967..4256
 FT /*tag- c
 FT exon


```

FT exon 4375..4334
FT /*tag= d
FT exon 4838..5075
FT /*tag= e
PN EP-297913-A.
PD 04-JAN-1989.
PF 30-JUN-1988; 306039.
PR 30-JUN-1987; US-068594.
PA (AMGE-) Amgen Inc.
PI Lin FK, Lu HS;
DR WPI: 89-009139/02.
P-PSDB: p93722.
PT New recombinant kallikrein polypeptide(s) and encoding DNA -
PS useful as vasodilators and for treating male infertility.
PT Table V; page 13; 43pp; English.
CC Lambda HK64a is a subclone produced from clones isolated from a CHA phage
CC bouane human fetal liver genomic library using monkey kallikrein MKK80a
CC clone DNA as a probe.
See also N93193.8.
Sequence 3406 BP; 974 A; 1864 C; 1309 G; 1259 T;

```

```

Query Match 100.0%; Score 10; DB 1; Length 5406;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTCCCGAG 10
DB 1223 GTCCTCCCGAG 1232

```

```

RESULT 6
ID 003143/c
AC 003143:
DE 10-MAR-1993 (revised)
DE 09-SEP-1990 (first entry)
DE Nucleotide sequence of cDNA from Aspergillus niger encoding glucose
DE oxidase
DE Glucose oxidase; Aspergillus niger; glucose detection;
DE KW glucose estimation.
OS Aspergillus niger.
FH Key Location/Qualifiers
FT misc_difference 2..2025
FT mat_peptide
/*tag= a
/*note="nucleotide sequence as given in Fig SB-1-5B-6"
98..1849
/*tag= b
32..97
/*tag= c
/product=peptide
PN WO8912675-A.
PD 28-DEC-1989.
PF 20-JUN-1988; U02696.
PR 21-JUN-1988; US-209530.
PR 19-JUN-1989; US-366377.
PA (ChIR-) Chiron Corp.
PI Rosenberg S;
DR WPI: 90-022535/03.
P-PSDB: R04724.
PT Recombinant polynucleotide(s) encoding glucose oxidase or a mutein -
PT and vectors contg. them, secreting glucose oxidase or
PT hyper-glycosylated analogues into medium
PS Disclosure; B-6; 108pp; English.
CC The cdna can also be used to isolate GO-encoding sequences from other
CC sources. Its encoded GO is used in desugaring eggs, removing oxygen from
CC moist food prodcs., beverages, flavours and hermetically-sealed food
CC packages, and in detection and estimation of glucose in industrial solns.
CC and body fluids, eg blood and urine. This sequence is the same as that
CC given in Fig 5B of the patent, but with the addition of one Guanine at
CC the start of the sequence.
Sequence 2024 BP; 450 A; 615 C; 515 G; 445 T;

```

```

Query Match 100.0%; Score 10; DB 1; Length 2024;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTCCCGAG 10
DB 1559 GTCCTCCCGAG 1550

```

```

RESULT 7
ID 004799/c
AC 004799:
DE 30-OCT-1990 (first entry)
DE Recombinant sequence encoding hepatitis B antigens.
DE Hepatitis B virus; vaccine; HBV; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT cds
/*tag= a
1..80
FT /*tag= b
87..634
FT /*tag= c
1521..2198
FT /*tag= c
EP-374869-A.
PD 27-JUN-1990.
PF 1-JAN-1989; 123526.
PR 22-DEC-1978; GB-049807.
PR 27-DEC-1978; GB-050039.
PR 01-NOV-1979; GB-037910.
PA (BIOJ) Biotogen Inc.
PI Murray K, Schaller HE;
DR WPI: 90-195067/26.
P-PSDB: R05634, R0535 & R06613.
DR Recombinant DNA encoding Hepatitis B virus polypeptide antigens -
PT used in detection of infection and in vaccine prodn.
PS Disclosure; English
CC Polypeptide fragments encoded by the DNA sequence are antigenic for
CC HBV and may be used as a vaccine or in detection. Peptides may be
CC cultured in a suitable bacterial host such as E.coli.
SQ Sequence 2743 BP; 638 A; 728 C; 579 G; 797 T; 1 Others;

```

```

Query Match 100.0%; Score 10; DB 1; Length 2743;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTCCCGAG 10
DB 279 GTCCTCCCGAG 270

```

```

RESULT 8
ID N81654
AC N81654:
DE 09-NOV-1990 (first entry)
DE LVDJ regions of human sperm-immobilising monoclonal antibody.
DE Anti-human sperm-immobilising monoclonal antibody; leader region;
DE variable region; joining region; constant region; sterility;
DE KW vaccine; contraceptive; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds
622..912
/*tag= a
/product=variable region
FT misc_feature
528..610
/*tag= b
/label=leader_region
FT misc_feature
622..912
/*tag= c
/label=variable_region
FT

```


FT	misc_feature	913..943	/*tag= d	
FT			/label-diversity-region	
FT		944..995		
FT	misc_feature		/*tag= e	
FT			/label=joining-region	
PN	J63126482-A.			
PD	30-MAY-1988.			
PF	15-NOV-1986; 272412.			
PR	15-NOV-1986; JP-272412.			
PR	(TOFU) TOA NENRYO KOGYO KK.			
PA	WPI; 88-187839/27.			
DR	P-PSDB; P81259.			
PT	Cell strain producing human sperm-immobilising monoclonal antibody - has at least V-gene in H chain coding genetic family and V gene in L chain of DNA originating from antibody of sterile woman.			
PS	Disclosure; 14pp; Japanese.			
CC	A 1.85 Kb Lymph segment is inserted into pSV29p together with a constant gamma 1 region. A cell strain, pref. a myeloma cell, can produce the Ab with immobilising value (SI50) of at least 5000, sperm agglutination value at least 1:1600 dilution, specifically reacting against human ejaculated sperm. The product can be used as a vaccine and contraceptive.			
CC	See also N81654-N81656.			
CC	Sequence 1071 BP; 250 A; 268 C; 281 G; 272 T;			
QY	Query Match	100.0%;	Score 10;	DB 1; Length 1071;
	Best Local Similarity	100.0%;	Pred. No. 1,1e+03;	
	Matches 10; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
DB	986 GTCCTCCAG 995			

Result	ID	Accession	Location/Qualifiers
10	011415/c	standard; cDNA; 15464 BP.	
11	011415	17-JUN-1991 (first entry)	
12	011415	Ryanodine receptor gene.	
13	011415	Malignant hyperthermia; hypermetabolic syndrome; inhalation;	
14	011415	anesthetics; probe; calcium release channel; sarcoplasmic;	
15	011415	reticulum; ss.	
16	011415	Homo sapiens.	
17	011415	Key	
18	011415	Key	
19	011415	Key	
20	011415	Key	
21	011415	Key	
22	011415	Key	
23	011415	Key	
24	011415	Key	
25	011415	Key	
26	011415	Key	
27	011415	Key	
28	011415	Key	
29	011415	Key	
30	011415	Key	
31	011415	Key	
32	011415	Key	
33	011415	Key	
34	011415	Key	
35	011415	Key	
36	011415	Key	
37	011415	Key	
38	011415	Key	
39	011415	Key	
40	011415	Key	
41	011415	Key	
42	011415	Key	
43	011415	Key	
44	011415	Key	
45	011415	Key	
46	011415	Key	
47	011415	Key	
48	011415	Key	
49	011415	Key	
50	011415	Key	
51	011415	Key	
52	011415	Key	
53	011415	Key	
54	011415	Key	
55	011415	Key	
56	011415	Key	
57	011415	Key	
58	011415	Key	
59	011415	Key	
60	011415	Key	
61	011415	Key	
62	011415	Key	
63	011415	Key	
64	011415	Key	
65	011415	Key	
66	011415	Key	
67	011415	Key	
68	011415	Key	
69	011415	Key	
70	011415	Key	
71	011415	Key	
72	011415	Key	
73	011415	Key	
74	011415	Key	
75	011415	Key	
76	011415	Key	
77	011415	Key	
78	011415	Key	
79	011415	Key	
80	011415	Key	
81	011415	Key	
82	011415	Key	
83	011415	Key	
84	011415	Key	
85	011415	Key	
86	011415	Key	
87	011415	Key	
88	011415	Key	
89	011415	Key	
90	011415	Key	
91	011415	Key	
92	011415	Key	
93	011415	Key	
94	011415	Key	
95	011415	Key	
96	011415	Key	
97	011415	Key	
98	011415	Key	
99	011415	Key	
100	011415	Key	

CC (HRR), a calcium release channel which spans the gap between the
 CC transverse tubule and the sarcoplasmic reticulum (SR) in muscle.
 CC The sustained muscle contraction in malignant hyperthermia (MH)
 CC may be caused by the release of calcium into the muscle cell cyto-
 CC plasm from the SR. This is due to a defect in the gene encoding
 CC HRR. Sequences from the gene can be used for diagnosis of MH
 CC using RFLP analysis (see tags e-m).
 SO Sequence 15464 BP; 3286 A; 4616 C; 4735 G; 2827 T;

Query Match 100.0%; Score 10; DB 1; Length 15464;
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 DB 2884 GTCCTCCAG 2875

.SULF 11
 ID N91627 standard; DNA; 1196 BP.

AC N91627:
 DT 19-MAR-1991 (first entry)
 DE Sequence of gamma-3 intron switch sequences for immunoglobulin
 DE genes.
 DE Immunoglobulin gene switch sequence; intron switch sequence; ss.
 OS Homo sapiens.

FT misc-feature 21..224 Location/Qualifiers
 FT /tag= a
 FT /label=sigma-gamma-3
 FT 225..1196
 FT /tag= b
 FT /label=sigma-gamma-4

PD 20-APR-1989.
 PF 14-OCT-1988; J01044.
 PR 09-JUL-1988; JP-170041.
 PA (EDUC-) Education Found Fujita (MITR MITR).
 PI Kuromawa Y, Away A, Jshizuka Y, Wakabayash T;
 DR WPT, 89-130043/17.
 PT Recombinant DNA switch sequences for immunoglobulin genes - enable
 PT specific control of prodn. of specific classes of antibody.
 PS Disclosure: Fig 9B; 79pp; Japanese.
 CC Sigma-gamma-3 and -4 are switch sequences for Cx genes upstream,
 CC encoding human IgX (X = IgG3, IgG1, Ig-gamma-E, IgM, IgG2, IgG4,
 CC IgE or IgA2). The DNA sequences are obtd. from suitable human cell
 CC line, e.g. marrow cells, and pref. ARA-10 or Daudi, using a probe
 CC for the IgH region, and cloned using a vector such as a plasmid of
 CC (pref.) lambda-phage.
 SO Sequence 1196 BP; 276 A; 312 C; 345 G; 263 T;

Query Match 100.0%; Score 10; DB 1; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 DB 880 GTCCTCCAG 889

RESULT 12
 Q11707/c
 ID Q11707 standard; DNA; 14928 BP.

AC Q11707:
 DT 25-JUL-1991 (first entry)
 DE Immunoglobulin D-region heavy chain major cluster.
 DE Lymphoid tumour diagnosis; human; Ig; diversity region; ss.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT misc-rna 776..862

FT /tag= a
 FT /label= Xp4 region
 FT 1924..1993
 FT /tag= b
 FT /label= A4 region
 FT 2885..2960
 FT /tag= c
 FT /label= K4 region
 FT 4734..4807
 FT /tag= d
 FT /label= N4 region
 FT 5238..5310
 FT /tag= e
 FT /label= M1 region
 FT 7921..8007
 FT /tag= f
 FT /label= LRI region
 FT 10447..10533
 FT /tag= g
 FT /label= Xp1 region
 FT 10631..10696
 FT /tag= h
 FT /label= Xp1 region
 FT 11522..11593
 FT /tag= i
 FT /label= A1 region
 FT 12478..12556
 FT /tag= j
 FT /label= K1 region
 FT 13963..14059
 FT /tag= k
 FT /label= N1 region
 FT 14490..14562
 FT /tag= l
 FT /label= M2 region
 FT 1575..1829
 FT /tag= m
 FT /label= Region I
 FT 3536..4394
 FT /tag= n
 FT /label= Region II
 FT 12686..13184
 FT /tag= o
 FT /label= Region III
 FT 13232..13247
 FT /tag= p
 FT /label= Region IV
 FT J03087185-A.
 PN 11-APR-1991.
 PD 19-DEC-1989; 329005.
 PF 19-DEC-1989; JP-319809.
 PR 19-DEC-1989; JP-154623.
 PR 19-JUN-1989; JP-329005.
 PA (MITR) MITSUI TOATSU CHEM INC.
 DR WPT; 91-152426/21.
 PT Human immunoglobulin gene - for diagnosis and monitoring of
 PT lymphoid tumour esp. in leukaemia
 PS Claim 2; Fig 2; 35pp; Japanese.
 CC The sequence is a 15kb fragment of the D(H) region of the human Ig
 CC gene (between D(LR4) and D(LR2)). See also Q11708.
 SO Sequence 14928 BP; 3308 A; 4995 C; 4066 G; 2559 T;

Query Match 100.0%; Score 10; DB 1; Length 14928;
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
 DB 6743 GTCCTCCAG 6734

RESULT 13


```

FT      /label= ATTA motif
FT      /note= "reduces mRNA stability"
FT      4240.4244
FT      misc_rna
FT      /tag= q
FT      /label= ATTA motif
FT      /note= "reduces mRNA stability"
FT      4249.4253
FT      misc_rna
FT      /tag= r
FT      /label= ATTA motif
FT      /note= "reduces mRNA stability"
FT      4280.4284
FT      misc_rna
FT      /tag= s
FT      /label= ATTA motif
FT      /note= "reduces mRNA stability"
FT      4357.4362
FT      poly_a_signal
FT      /tag= t
FT      4357.4362
FT      WO910738-A.
PN      25-JUL-1991.
PD      08-JAN-1991; US-462158.
DR      (RESE ) RES CORP TECHN INC.
PI      Vansnick J, Uytendhoeve C, Simpson RJ;
DR      WPI; 91-238026/32.
DR      P-PDB; R13217.
PT      Nucleic acid encoding for P40 T-cell growth factor - for patients
PT      with AIDS or compromised immune systems, also allows increased
PT      prodn. of other cytokine(s).
PS      Claim 2; Fig 14; 122pp; English.
CC      The sequence was obt'd. from a clone, lambda H40.3a1, isolated from
CC      a genomic library using a murine P40 CDNA clone as a probe. The
CC      sequence, or the cDNA sequence obt'd. using the sequence, can be
CC      inserted into a vector for expression of P40 in a host organism.
CC      The protein is useful for stimulating the proliferation of certain
CC      subsets of T helper cells e.g. in AIDS patients or immune
CC      compromised patients.
CC      See also Q12760, and Q13660-Q13663.
SQ      Sequence 4382 BP, 1210 A; 922 C; 904 G; 1346 T;

```

```

Query Match      100.0%; Score 10; DB 1; Length 4382;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GTCCTCCCGAG 10
      |||||
DB      367 GTCCTCCCGAG 376

```

Search completed: September 18, 1999, 00:33:39
Job time: 18979 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:06 : Search time 2825.05 Seconds

(without alignments)
6.982 Million cell updates/sec

Title: US-08-956-518a-93
Perfect score: 10
Sequence: 1 GTCTCCAG 10

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	10	100.0	200	20	D20867	D20867 HUMS01848
C 2	10	100.0	161	20	D25774	D25774 HUMS04142
C 3	10	100.0	166	20	D28621	D28621 MDS0601 MO
C 4	10	100.0	424	20	D32336	D32336 CELK01184R
C 5	10	100.0	360	20	D35681	D35681 CELK02487F
C 6	10	100.0	360	20	D36081	D36081 CELK02883F
C 7	10	100.0	418	20	D37701	D37701 CELK0808DF
C 8	10	100.0	391	20	D41072	D41072 RICS3337A R
C 9	10	100.0	324	20	F02878	F02878 HSC1ED092 n
C 10	10	100.0	268	20	F03210	F03210 HSC1ED082 n
C 11	10	100.0	521	20	L26678	L26678 MDSF080A 1a
C 12	10	100.0	401	20	L26715	L26715 MDS167A 1am
C 13	10	100.0	309	20	L34981	L34981 HUMX20EST H
C 14	10	100.0	397	20	M78691	M78691 EST00839 H1
C 15	10	100.0	338	20	M85776	M85776 EST02297 Fe
C 16	10	100.0	411	20	M89223	M89223 CEL1868 Chr
C 17	10	100.0	340	20	T00993	T00993 WEST01714 E
C 18	10	100.0	236	20	T00994	T00994 WEST01715 E
C 19	10	100.0	352	20	T03019	T03019 FB20D6 Feta
C 20	10	100.0	213	20	T03520	T03520 IB403 Infan
C 21	10	100.0	199	20	T03860	T03860 seq2338 4HF
C 22	10	100.0	238	20	T04845	T04845 EST02733 Fe
C 23	10	100.0	289	20	T06308	T06308 EST04197 Fe
C 24	10	100.0	374	20	T06613	T06613 EST04502 Fe
C 25	10	100.0	343	20	T07835	T07835 EST05725 Fe
C 26	10	100.0	379	20	T10329	T10329 seq1249 b4H
C 27	10	100.0	1069	20	T12278	T12278 EST0003 Sub
C 28	10	100.0	146	20	T12345	T12345 A044F Heart
C 29	10	100.0	397	20	T15603	T15603 IB1616 Infa
C 30	10	100.0	250	20	T16832	T16832 NIB1910-5R
C 31	10	100.0	450	20	T18269	T18269 5C01905-t7
C 32	10	100.0	408	20	T19397	T19397 f02001s Tes
C 33	10	100.0	397	20	T24183	T24183 CTS171 1amB
C 34	10	100.0	363	20	T27108	T27108 NIBT290C05R
C 35	10	100.0	340	20	T27693	T27693 EST12419 Hu
C 36	10	100.0	213	20	T28192	T28192 EST131471 Hu
C 37	10	100.0	253	20	T29572	T29572 EST185052 Hu
C 38	10	100.0	271	20	T30400	T30400 EST15995 Hu
C 39	10	100.0	269	20	T30791	T30791 EST12689 Hu
C 40	10	100.0	448	20	T33051	T33051 EST156484 Hu
C 41	10	100.0	364	20	T33660	T33660 EST158625 Hu
C 42	10	100.0	202	20	T35723	T35723 EST190125 Hu
C 43	10	100.0	345	20	T36837	T36837 EST101849 S
C 44	10	100.0	307	20	T37940	T37940 EST103255 S
C 45	10	100.0	490	54	HSM009637	A1044787 Homo sapi

ALIGNMENTS

RESULT 1
D20867/c
LOCUS HUMS01848 Human promyelocyte Homo sapiens cDNA clone mp224 3',
DEFINITION mRNA sequence.
ACCESSION D20867
NID 9504687
VERSION D20867.1 GI:504687

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 200)
Miyakawa, K., Matsubara, K., Fukushima, A., Yoshii, J., and Okubo, K.
JOURNAL Chromosomal assignments of 3'-directed partial cDNA sequences
MEDLINE representing novel genes expressed in granulocytoid cells
95137584
Genomics 23, 379-389 (1994)

REFERENCE
AUTHORS 2 (bases 1 to 200)
Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,
Yoshinari, H., Arimoto, J., and Matsubara, K.
TITLE gene expression of human promyelocytic cell line HL60 before and
after induction of differentiation. A new application of 3'-directed
cDNA sequencing
JOURNAL Unpublished (1993)
COMMENT

FEATURES
SOURCE 1..200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6"
/clone="mp2234"
/clone_1lb="Human promyelocyte"
/note="Female, adult, cell_line = HL60, cell_type =
promyelocyte."

BASE COUNT 55 a 58 c 40 g 42 t 5 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 200;
Best Local Similarity 100.0%; Pred. No. 8.4e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
Db 27 GTCCTCCAG 18

RESULT 2
5774/c 161 bp mRNA EST 30-NOV-1995
-JUS HUMS04142 Human colon mucosa Homo sapiens cDNA clone cm0231 3',
DEFINITION mRNA sequence.
ACCESSION D25774
NID 9500461
VERSION D25774.1 GI:500461
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 161)
Okubo, K., Yoshii, J., Yokouchi, H., Kamayama, M., and Matsubara, K.
JOURNAL Global analysis of gene expression in colon mucosa: a large scale
random cDNA sequencing analysis
COMMENT Unpublished (1994)
Contact: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H., and Matsubara, K.
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers
1..161
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 47 a 31 c 37 g 46 t
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 161;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
Db 26 GTCCTCCAG 17

RESULT 3
D28621 166 bp mRNA EST 07-OCT-1996
LOCUS M0586G01 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
DEFINITION clone 86G01, mRNA sequence.
ACCESSION D28621
NID 9518938
VERSION D28621.1 GI:618938
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 166)
Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T., and Shimada, K.
AUTHORS A survey of genes expressed in undifferentiated mouse embryonal
TITLE carcinoma F9 cells: characterization of low-abundance mRNAs
JOURNAL J. Biochem. 116, 128-139 (1994)
MEDLINE 95096008
COMMENT

CONTACT: Kazunori Shimada
Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.
Location/Qualifiers
1..166
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="86G01"
/clone_1lb="mouse embryonal carcinoma cell line F9"

BASE COUNT 33 a 44 c 43 g 44 t 2 others
ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 166;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCAG 10
Db 20 GTCCTCCAG 11

RESULT 4
D32336 424 bp mRNA EST 05-AUG-1994
LOCUS CELK011BAR Yujii Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk11b4 3', mRNA sequence.
ACCESSION D32336
NID 9522953
VERSION D32336.1 GI:522953
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdillia; Rhabdillida;

REFERENCE 1 (bases 1 to 424)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.j.
 Location/Qualifiers

FEATURES
 source 1..424
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK11b4"
 /clone_1ib="Yuji Kohara unpublished CDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 150 a 78 c 83 g 112 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 424;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 337 GTCCTCCCGAG 346

RESULT 5
 D35681/c 360 bp mRNA EST 08-AUG-1994
 LOCUS CELK02487F Yuji Kohara unpublished CDNA Caenorhabditis elegans cDNA
 DEFINITION clone yk24b 5', mRNA sequence.
 ACCESSION D35681
 NID 9527131
 VERSION D35681.1 GI:527131
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 360)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.j.
 Location/Qualifiers

FEATURES
 source 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK24b7"
 /clone_1ib="Yuji Kohara unpublished CDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 144 a 49 c 88 g 78 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;
 Best Local Similarity 100.0%; Pred. No. 9.7e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 202 GTCCTCCCGAG 193

RESULT 6
 D36081/c 360 bp mRNA EST 08-AUG-1994
 LOCUS CELK028H3F Yuji Kohara unpublished CDNA Caenorhabditis elegans cDNA
 DEFINITION clone yk28h3 5', mRNA sequence.
 ACCESSION D36081
 NID 9528019
 VERSION D36081.1 GI:528019
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 360)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.j.
 Location/Qualifiers

FEATURES
 source 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK28h3"
 /clone_1ib="Yuji Kohara unpublished CDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"
 BASE COUNT 96 a 86 c 124 g 51 t 3 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 360;
 Best Local Similarity 100.0%; Pred. No. 9.7e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 248 GTCCTCCCGAG 239

RESULT 7
 D37701/c 418 bp mRNA EST 08-AUG-1994
 LOCUS CELK008D8F Yuji Kohara unpublished CDNA Caenorhabditis elegans cDNA
 DEFINITION clone yk8d 5', mRNA sequence.
 ACCESSION D37701
 NID 9526082
 VERSION D37701.1 GI:526082
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 418)

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.j.
 Location/Qualifiers

AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

FEATURES
 source
 Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1.418
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK868"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /note="dev-stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"
 /tissue_type="whole animal"

BASE COUNT 104 a 94 c 99 g 121 t
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCCTCCCG 10
 |||||||||
 Db 395 GTCCTCCCG 386

RESULT 8
 LOCUS D41072 391 bp mRNA EST 15-NOV-1994
 DEFINITION R1C33337A Rice shoot Oryza sativa cDNA, mRNA sequence.
 ACCESSION D41072
 NID 9571561
 VERSION D41072.1 GI:571561
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 391)
 Sasaki, T., Miyao, A. and Yamamoto, K.
 Rice cDNA from callus 1995
 Unpublished (1995)
 COMMENT
 Contact: Takuji Sasaki
 National Institute of Agricultural Resources
 Rice Genome Research Program
 2-1-2 Kammonadal, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@dbj.nig.ac.jp.
 Location/Qualifiers
 1.391
 /organism="Oryza sativa"
 /strain="Nipponbare sub-species Japonica"
 /db_xref="taxon:4530"
 /clone_lib="Rice shoot"
 /note="Etiolated shoot (8 days old)"

BASE COUNT 72 a 140 c 119 g 54 t 6 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 391;

Best Local Similarity 100.0%; Pred. No. 9.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCCTCCCG 10
 |||||||||
 Db 67 GTCCTCCCG 76

RESULT 9
 LOCUS F02878/c 324 bp mRNA EST 02-FEB-1995
 DEFINITION HSC1E092 normalized infant brain cDNA Homo sapiens cDNA clone
 c-1ed09 3', mRNA sequence.
 F02878
 ACCESSION F02878
 NID 9646435
 VERSION F02878.1 GI:646435
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 324)
 Aulicay, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
 Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B.,
 Lorenzo, F., Mitchell, H., Maringe-Samson, R., Pletu, G., Pouliot, Y.,
 Sebastiani, K., Katchis, C. and Tessier, A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534

JOURNAL
 MEDLINE
 COMMENT
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 3316078698
 Email: genexpress@genethon.fr
 Single read. removed at sequence 5' end
 Genexpress_library_id: C; Genexpress_sequence_id: alc-1ed09
 Seq primer: (-21)M13 universal.

FEATURES
 source
 Location/Qualifiers
 1.324
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="c-1ed09"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lambda B. Site: 1, HindIII;
 Site: 2, NotI; sex=Female; dev-stage=3 months old;
 isolate=muscular atrophy patient; tissue_type=total
 brain; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambda B. vector. Clone library from B. Soares, psychiatry
 Dept. Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S. in press"

BASE COUNT 67 a 84 c 119 g 53 t 1 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 324;
 Best Local Similarity 100.0%; Pred. No. 9.4e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCG 10
 |||||||||
 Db 228 GTCCTCCCG 219

RESULT 10
 F03210/c

LOCUS F03210 268 bp mRNA EST 02-FEB-1995
 DEFINITION HSC1P6082 normalized infant brain CDNA Homo sapiens CDNA clone
 c-1P608 3', mRNA sequence.
 ACCESSION F03210
 NID 9646767
 VERSION F03210.1 GI:646767
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 268)
 Aufray,C., Benar,G., Bois,F., Bouchier,C., da Silva,C.,
 Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B.,
 Lorenzo,F., Mitchell,H., Marigault-Sanson,R., Pletu,G., Pouillot,Y.,
 Sebastiani-Kabackich,C. and Tessier,A.
 IMAGE: molecular integration of the analysis of the human genome
 and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534

TITLE
 JOURNAL
 MEDLINE
 JUMENT

CONTACT: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read, removed at sequence 5' end
 Genexpress_library_id: C; Genexpress_sequence_id: alc-1P608
 Seq primer: (-21)M13.universal.
 Location/Qualifiers
 1. 268
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="c-1P608"
 /clone_1id="normalized infant brain CDNA"
 /sex="Female"
 /tissue-type="total brain"
 /dev_stage="3 months old"
 /note="Organ: Brain; Vector: lambda BA; Site_1: HindIII;
 Site_2: NotI; sex:Female; dev_stage=3 months old;
 isolate=muscular atrophy patient; tissue-type=total
 brain; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambda BA vector. Clone library from B. Soares, Psychiatry
 Dept. Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S. in press"

BASE COUNT 74 a 53 c 71 g 67 t 3 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 268;
 Best Local Similarity 100.0%; Pred. NO. 9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 119 GTCTCCCGAG 110

RESULT 11
 L26678 521 bp mRNA EST 01-MAR-1995
 LOCUS M05F080A lambda unizap male testis Mus musculus CDNA clone F080.
 DEFINITION mRNA sequence.
 ACCESSION L26678
 NID 9437469
 VERSION L26678.1 GI:437469
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 521)
 AUTHORS Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 TITLE Analysis of cDNA sequences from mouse testis
 JOURNAL Mamm. Genome 5, 557-565 (1994)
 MEDLINE 95093181
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 Location/Qualifiers
 1. 521
 /organism="Mus musculus"
 /strain="Swiss"
 /db_xref="taxon:10090"
 /clone="F080"
 /clone_1id="lambda unizap male testis"
 /note="Vector: lambda unizap; lambda unizap library from
 mall mouse testis"

BASE COUNT 131 a 127 c 129 g 132 t 2 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 521;
 Best Local Similarity 100.0%; Pred. NO. 1.1e+04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 216 GTCTCCCGAG 225

RESULT 12
 L26715 401 bp mRNA EST 01-MAR-1995
 LOCUS M05167A lambda unizap male testis Mus musculus CDNA clone F167.
 DEFINITION mRNA sequence.
 ACCESSION L26715
 NID 9437393
 VERSION L26715.1 GI:437393
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 401)
 Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 TITLE Analysis of cDNA sequences from mouse testis
 JOURNAL Mamm. Genome 5, 557-565 (1994)
 MEDLINE 95093181
 COMMENT Contact: Kerr,S.M., Vambrie,S., McKay,S.J. and Cooke,H.J.
 Location/Qualifiers
 1. 401
 /organism="Mus musculus"
 /strain="Swiss"
 /db_xref="taxon:10090"
 /clone="F167"
 /clone_1id="lambda unizap male testis"
 /note="Vector: lambda unizap; lambda unizap library from
 mall mouse testis"

BASE COUNT 113 a 87 c 124 g 71 t 6 others
 ORIGIN

Query Match 100.0%; Score 10; DB 20; Length 401;
 Best Local Similarity 100.0%; Pred. NO. 9.9e+03;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCGAG 10
 |||||||||
 Db 274 GTCTCCCGAG 265

RESULT 13

L34981
 LOCUS L34981 309 bp mRNA EST
 DEFINITION HUM20EST Human brain striatum Homo sapiens cDNA clone Y20, mRNA
 sequence.
 ACCESSION L34981
 NID 9522224
 VERSION L34981.1 GI:5222224
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Jou, Y.-S., Gould, R.D. and Myers, R.M.
 TITLE Localization of the alpha 2-macroglobulin receptor-associated protein 1 gene (LRPAP1) and other gene fragments to human chromosome 4p16.3 by direct CDNA selection
 JOURNAL Genomics 24, 410-413 (1994)
 MEDLINE 95213044
 COMMENT
 FEATURES
 source Contact: Jou, Y.-S., Gould, R.D. and Myers, R.M.
 Location/Qualifiers
 1. 309
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="4p16.3"
 /clone="Y20"
 /clone_lib="Human brain striatum"
 /note="tissue_type = brain striatum"
 BASE COUNT 38 a 84 c 87 g 94 t 6 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 10; DB 20; Length 309;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCG 10
 Db 1 GTCCTCCCG 10

RESULT 14
 M78691/c 397 bp mRNA EST 26-MAY-1992
 LOCUS M78691/c
 DEFINITION EST00839 Hippocampus, Striatum (cat. #936205) Homo sapiens cDNA
 clone HRCMC15, mRNA sequence.
 ACCESSION M78691
 NID 9273006
 VERSION M78691.1 GI:273006
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
 1 (bases 1 to 397)
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT
 FEATURES
 source Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1. 397
 /organism="Homo sapiens"

/db_xref="ATCC (inhost):78098"
 /db_xref="GDB:DS1679E"
 /db_xref="taxon:9606"
 /clone="HRCMC15"
 /clone_lib="Hippocampus, Striatum (cat. #936205)"
 /note="Vector: LambdaZAP-II; Female, 2 years; oligo-dT +
 random primed cDNA synthesis; LambdaZAP-II vector, 1.0kb
 average insert size."
 BASE COUNT 104 a 106 c 96 g 88 t 3 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 10; DB 20; Length 397;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTCCCG 10
 Db 289 GTCCTCCCG 280

RESULT 15
 M85776/c 338 bp mRNA EST 11-JAN-1995
 LOCUS M85776
 DEFINITION EST02297 Fetal brain, Striatum (cat#936206) Homo sapiens cDNA
 clone HRCMC15, mRNA sequence.
 ACCESSION M85776
 NID 9274425
 VERSION M85776.1 GI:274425
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
 1 (bases 1 to 338)
 Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C.
 TITLE Sequence identification of 2,375 human brain genes
 JOURNAL Nature 355, 632-634 (1992)
 MEDLINE 92168112
 COMMENT
 REFERENCE
 AUTHORS Durdin, A.S.
 TITLE EST mapping data
 JOURNAL Unpublished (1996)
 COMMENT
 FEATURES
 source Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers
 1. 338
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):81353"
 /db_xref="GDB:DS581E"
 /db_xref="taxon:9606"
 /map="8"
 /clone="HRCMC15"
 /clone_lib="Fetal brain, Striatum (cat#936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; LambdaZAP-II
 vector, 1.0kb average insert size."
 BASE COUNT 85 a 65 c 94 g 92 t 2 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 10; DB 20; Length 338;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SEP 19 10:27:03 1999

us-08-956-518a-93.rst

Page 7

OY	1	GTCTCCCGAG	10
DB	44	GTCTCCCGAG	35

Search completed: September 17, 1999, 21:28:07
Job time: 14271 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:01:46 ; Search time 1962.17 Seconds

(without alignments)
740.711 million cell updates/sec

Title: US-08-956-518a-94

Sequence: 1 AGAAGCAGGAGGAGGAGTAG.....CTCTGACAGGTAAGCCAC 457

Scoring table: IDENTITY_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:.*
1: gb_dal:.*
2: gb_dal:.*
3: gb_om:.*
4: gb_ov:.*
5: gb_pat:.*
6: gb_ph:.*
7: gb_pl:.*
8: gb_pl:.*
9: gb_pl:.*
10: gb_pl:.*
11: gb_pl:.*
12: gb_pl:.*
13: gb_pl:.*
14: gb_pl:.*
15: gb_pl:.*
16: gb_pl:.*
17: gb_pl:.*
18: gb_pl:.*
19: gb_pl:.*
20: gb_pl:.*
21: gb_pl:.*
22: gb_pl:.*
23: gb_pl:.*
24: gb_pl:.*
25: gb_pl:.*
26: gb_pl:.*
27: gb_pl:.*
28: gb_pl:.*
29: gb_pl:.*
30: gb_pl:.*
31: gb_pl:.*
32: gb_pl:.*
33: gb_pl:.*
34: gb_pl:.*
35: gb_pl:.*
36: gb_pl:.*
37: gb_pl:.*
38: gb_pl:.*
39: gb_pl:.*
40: gb_pl:.*
41: gb_pl:.*
42: gb_pl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	392	85.8	392 11	AF029837 Homo sapi

2	127	27.8	1876	10	HSU62436	U62436 Human nicot
3	121.6	26.6	2087	10	HSAR7A	X70297 H. sapiens m
4	73.6	16.1	168919	35	AC007686	AC007686 Homo sapi
5	70	15.3	342379	34	CEY39B6	Z55399 Caenorhabdi
6	68	14.9	1072	17	HS21E	M29384 Herpes simp
7	65.8	14.4	1560	10	HS21E	Z84721 Human DNA s
8	63.2	13.8	43058	10	HS21E	J00134 Human alpha
9	63.2	13.8	2685	10	HUMBA3	Z86099 Herpes simp
10	63	13.8	154746	17	HSV2HG52	U40583 Human alpha
11	62	13.6	154746	17	HSV2HG52	AT049628 Streptomy
12	62	13.6	1977	10	HSU40583	AD006684 Homo sapi
13	61.6	13.5	38532	10	SCB94	Y07651 H. sapiens u
14	61.6	13.5	41936	10	CH19R30879	Y11354 H. sapiens m
15	61.6	13.5	14440	10	HSUSP2	AC006443 Homo sapi
16	61.4	13.4	3252	10	HSUSP2	AC007225 Homo sapi
17	61.4	13.4	21636	11	HSUSP2	Z12168 C. famillari
18	61.2	13.4	248287	35	CEGTPBPAM	AB016897 Homo sapi
19	61	13.3	1927	3	CEGTPBPAM	AB016897 Homo sapi
20	61	13.3	557	4	XELRGE12	AL009178 Human DNA
21	61	13.3	3924	4	XELRGE12	AL009178 Human DNA
22	61	13.3	7634	4	XELRGE12	AL009178 Human DNA
23	61	13.3	8133	4	XELRGE12	AL009178 Human DNA
24	59	12.9	634	14	G40486	AF030453 Homo sapi
25	58.2	12.7	163974	9	AB011399	G40486 26971 Zebra
26	58.2	12.7	331211	9	AB011399	AB011399 Homo sapi
27	58.2	12.7	147971	10	HS431P23	AB016897 Homo sapi
28	58.2	12.7	179436	35	AC003066	AL009178 Human DNA
29	57.8	12.6	181884	34	HS35C18	AC003066 Mus muscu
30	57.6	12.6	240229	35	AC003059	AL022327 Homo sapi
31	57.2	12.5	1027	12	MMIGFITE2	AC003059 Mus muscu
32	57.2	12.5	267925	35	AC007298	X71919 H. musculus
33	57	12.5	557	4	XELRGE12	K01371 X. laevis oo
34	57	12.5	3924	4	XELRGE12	J00999 X. laevis ex
35	57	12.5	7634	4	XELRGE12	X59734 X. laevis ex
36	57	12.5	8133	4	XELRGE12	X02995 Xenopus lae
37	56.8	12.4	204263	35	AC007224	AF030453 Homo sapi
38	56.4	12.3	39213	11	AC003047	X02995 Xenopus lae
39	56.4	12.3	152261	17	HEICG	AC007224 Homo sapi
40	56.4	12.3	152261	17	HEICG	AC003047 Homo sapi
41	56.4	12.3	152261	17	HEICG	X14112 Herpes simp
42	56.4	12.3	26245	17	HSUSV1G3	X14112 Herpes simp
43	56.4	12.3	26245	17	HSUSV1G3	X06461 Herpes simp
44	56.4	12.3	26245	17	HSUSV1G3	L00036 HSV1 (stra)
45	56.2	12.3	1766	17	HSUSV1G3	L00036 HSV1 (stra)

ALIGNMENTS

RESULT 1

AF029837

LOCUS

DEFINITION

REGION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

AF029837 392 bp DNA PRI 16-DEC-1998
Homo sapiens alpha 7 neuronal nicotinic receptor gene, promoter
region.
AF029837
G3757792
AF029837.1 GI:3757792
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Buiting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)
9900837
2 (bases 1 to 392)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.

TITLE		Direct Submission	
JOURNAL		Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA	
FEATURES		Location/Qualifiers	
SOURCE		1. 392	
PROMOTER		/organism="Homo sapiens"	
BASE COUNT		/db_xref="taxon:9606"	
ORIGIN		/chromosome="15"	
		/map="15q14; between D15S1031 and D15S1360"	
		/clone="CHRNA7-P"	
		1. 392	
		51 a 122 c 180 g 39 t	
Query Match		85.8%; Score 392; DB 11; Length 392;	
Best Local Similarity		100.0%; Pred. No. 2.7e-46;	
Matches 392; Conservative		0; Mismatches 0; Indels 0; Gaps	
QY	1	AGAACCGAAGGAGAGGTAGAGCTGGCTTTGGGCAACCCCTGGCTGGCCAGAGGCGG	60
Db	1	AGAACCGAAGGAGAGGTAGAGCTGGCTTTGGGCAACCCCTGGCTGGCCAGAGGCGG	60
QY	61	AGGCGGAGAGCCCGCTGGTGAGACACCTGGGGGCTGGAGTCCCGGAGCCGATCCAGCGCC	120
Db	61	AGGCGGAGAGCCCGCTGGTGAGACACCTGGGGGCTGGAGTCCCGGAGCCGATCCAGCGCC	120
QY	121	GGGAGTACTCCCGCTCAACACCTGGGGCTGAGTTCCTGGGTGGCCGCGAGACGCTGG	180
Db	121	GGGAGTACTCCCGCTCAACACCTGGGGCTGAGTTCCTGGGTGGCCGCGAGACGCTGG	180
QY	181	CCCGGCTGAGAGGATGGCGGGGCGGGGACGCGGGGCGGGGCGGGGCTCTCACTGGAG	240
Db	181	CCCGGCTGAGAGGATGGCGGGGCGGGGACGCGGGGCGGGGCGGGGCTCTCACTGGAG	240
QY	241	AGCGCGCGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAGGCGGCGAGCC	300
Db	241	AGCGCGCGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAGGCGGCGAGCC	300
QY	301	GAGGGGAGAGGTGCTCTGTGGCGGACAGGGGACGCGGGGCGAGCGAGACGTGGAG	360
Db	301	GAGGGGAGAGGTGCTCTGTGGCGGACAGGGGACGCGGGGCGAGCGAGACGTGGAG	360
QY	361	CGCGCGGGCTGCTGTCAGACTCCGGGACTCAAC	392
Db	361	CGCGCGGGCTGCTGTCAGACTCCGGGACTCAAC	392
RESULT	2		
LOCUS	HSU62436	1876 bp	11-JAN-1997
DEFINITION	Human nicotinic acetylcholine receptor alpha7 subunit precursor,		
ACCESSION	U62436		
NID	91458119		
VERSION	062436.1		
KEYWORDS	GI:1458119		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homindaes; Homo.		
TITLE	1 (bases 1 to 1876)		
JOURNAL	Elliot,R.J., Ellis,S.B., Bercchan,K.J., Urrutia,A.,		
MEDLINE	Chaver-Noriega,L.E., Johnson,E.C., Velicelebi,G. and Harpold,M.M.		
REFERENCE	Comparative structure of human neuronal alpha 2-alpha 7 and beta		
AUTHORS	2-beta 4 nicotinic acetylcholine receptor subunits and functional		
TITLE	expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and		
	beta 4 subunits		
	J. Mol. Neurosci. 7 (3), 217-228 (1996)		
	2 (bases 1 to 1876)		
	Elliot,R.J.		
	Direct Submission		

JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBIA Neurosciences, Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA

FEATURES
SOURCE
Location/Qualifiers
1. .1876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Kc1alpha7.3"
/clone_id="SIBIA.lamba gtl1 library #3 (M. Williams)"
/cell_line="IMR32"
/cell_type="neuroblastoma"
1. .72
73. .138
73. .1581
/codon_start=1
/product="nicotinic acetylcholine receptor alpha7 subunit precursor"
/protein_id="AAB40114.1"
/db_xref="PID:91458120"
/db_xref="GI:1458120"
/translation="MRCSPGCVWALALASLLHVSLOEFOKLYKELVKNINPLERP
AMDSQPLTVYFSLIQIMDVDEKNQVLTINIMLOMSMDTHYLIQWNSVEYGVTR
PDQGIKRPDILLVNSADERDATEHTVNLVNSGHCQILPGIKRSCEYIDVRNFP
VORCKTFKGSWSYSGSLDIOMOPADISGYIPNEMDPLVGIKRSERFCEKPEY
DMETVMMRRRTLYVGINLIPCVLISALILVPLPADSEKRSIGITVLSLTFVE
LVAELIPATSDSVPFLAQYFASMTIVGSLVYTVLYQTHHDDPDGKRPKTRV
LLNCAWFLRMKRPGEDEVKRPACQKHQRCSLASVEASVAPPPASNSGLLYIGRGI
DGWACVPTPDGSGVCGMACSPTHDEHLHSGCPEDPDPAKILLEVRYIANFRRC
DESAVCSSEKMFACVVDRLCLMAFSVFTICITIGILMSAPNFEVASKDBA"
139. .1578
/product="nicotinic acetylcholine receptor alpha7 subunit
1582. .1876

BASE COUNT 369 a 553 c 531 g 423 t

ORIGIN

Query Match 27.8%; Score 127; DB 10; Length 1876;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 321 GGGCGAGGCGGAGGCGGAGGCGAGAGCGTGAGGAGCGCGGCGTGTGCTGCTCAGCT 380
Db 1 GGGCGAGGCGGAGGCGGAGGCGAGAGCGTGAGGAGCGCGGCGTGTGCTCAGCT 60
Y 381 CCGGAGCTCAACATGCGTGTGCGCGGAGAGCGTGTGCTGCGCTGGCGCGTGTGCTC 440
Db 61 CCGGAGCTCAACATGCGTGTGCGCGGAGAGCGTGTGCTGCGCTGGCGCGTGTGCTC 120
Y 441 CTGCACG 447
|||||
Db 121 CTGCACG 127

RESULT 3
LOCUS HSNARA7A 2087 bp mRNA PRI 01-JUN-1994
DEFINITION H.sapiens mRNA for neuronal nicotinic acetylcholine receptor
alpha-7 subunit.
ACCESSION X70297
NID 9486606
KEYWORDS X70297.1 GI:496606
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2087)
AUTHORS Peng,X., Katz,M., Gerzanich,V., Anand,R. and Lindstrom,J.
TITLE Human alpha 7 acetylcholine receptor: cloning of the alpha 7
subunit from the SH-SY5Y cell line and determination of
pharmacological properties of native receptors and functional alpha
7 homomers expressed in Xenopus oocytes
JOURNAL Pharmacol. 45 (3), 546-554 (1994)
MEDLINE 94195283

REFERENCE 2 (bases 1 to 2087)
 AUTHORS Katz, M.
 TITLE Direct Submission
 JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235 Stemmler Hall, Philadelphia, PA 19104, USA
 FEATURES
 source
 1. 2087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /cell_type="neuroblastoma"
 /cell_line="SHSY-5Y"
 /clone_id="lambda zap II"
 /clone="SHSY3"
 104..175
 /product="neuronal nicotinic acetylcholine receptor
 alpha-7 subunit"
 104..1612
 /product="neuronal nicotinic acetylcholine receptor
 alpha-7 subunit"
 /codon_start=1
 /product="neuronal nicotinic acetylcholine receptor
 alpha-7 subunit"
 /protein_id="CAA49778.1"
 /db_xref="PIR:P496607"
 /db_xref="GI:496607"
 /db_xref="SWISS-PROT:P36544"
 /translation="MRCSPGVVWLGAAALHYSLOGEPRKLYKELVKNYPLNPV
 ANDSPLIVFISLQIMDVDEKNVLTNIMLQMSWIDHYLQMWSEPGVTVRF
 PQGLWPKPILLYNSADERDAFHNVLVNSSHCQYLPPIGFKSCYIDVMPFED
 VOHCKLKFGSWSGSLDLOMOEADISGIIPIGNGEWDLVIGPKRSEKCEKPEP
 DVTFTVTRRRRTLYGILNLLPCVLLSALALVELLPADSGEKISIGITVLSLTVEM
 LLYAEIMPATSDSVPLIAQYFASTMIVTIVSVYTVIVQYHNDGDKMKPMTYVI
 LNNKMWPLRMKRGEDKRPAPACQHQKRCISLASVEMSAVGPAPASNNGLTIGFEGGL
 DEHYCVPTPDGSGVCGRMACSPHDEHLHGGQPPGADPDLAILEVERIANRFGCQ
 DSEAVCSMKRAACVYDRCLMAFVSFTIICITIGISMNPNEVAVSRDFA"
 mat_peptide 441 a 601 c 573 g 472 t
 BASE COUNT 176..1609
 ORIGIN
 Query Match 26.6%; Score 121.6; DB 10; Length 2087;
 Best Local Similarity 96.1%; Pred. No. 3.2e-09;
 Matches 146; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
 QY 296 GAGCCGAGCGCCGAGGCTCTGTGCGCCGAGCGCGCGGAGACCGGAGCG 355
 |||||||
 Db 9 GAGCCGAGCGCGGAGGCTCTGTGCGCCGAGCGCGCGGAGACCGGAGCG 66
 |||||||
 QY 356 TGGAGCGCGCGGCTGCTGCTGCGGAGCAACATGCGGCTGCGCGGAGCGCT 415
 |||||||
 Db 67 TGGAGCGCGCGGCTGCTGCTGCGGAGCAACATGCGGCTGCGCGGAGCGCT 126
 |||||||
 QY 416 CTGGCTGGCGCTGCGCGGCTGCTGCTGCGAGC 447
 |||||||
 Db 127 CTGGCTGGCGCTGCGCGGCTGCTGCTGCGAGC 158
 |||||||
 RESULT 4
 AC007686 168919 bp DNA HTG 30-MAY-1999
 LOCUS Homo sapiens chromosome 14 clone BAC 2289B16 map 14q24.3, LOW-PASS
 DEFINITION SEQUENCE SAMPLING.
 AC007686
 NID 94927297
 VERSION AC007686.1 GI:4927297
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 168919)
 REFERENCE
 AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,
 Shaffer, T. and Hood, L.
 Sequencing of human chromosome 14q24.3 region
 2 (bases 1 to 168919)
 Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,
 Shaffer, T. and Hood, L.
 Direct Submission
 Submitted (30-MAY-1999) Multimegabase Sequencing Center, University
 of Washington, PO Box 357730, Seattle, WA 98195, USA
 NOTE: his record contains 192 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.
 1
 965: contig of 965 bp in length
 gap of unknown length
 966 1866: contig of 901 bp in length
 gap of unknown length
 1867 2747: contig of 881 bp in length
 gap of unknown length
 2748 3608: contig of 861 bp in length
 gap of unknown length
 3609 4462: contig of 854 bp in length
 gap of unknown length
 4463 5345: contig of 883 bp in length
 gap of unknown length
 5346 6187: contig of 842 bp in length
 gap of unknown length
 6188 7072: contig of 885 bp in length
 gap of unknown length
 7073 7927: contig of 855 bp in length
 gap of unknown length
 7928 8797: contig of 870 bp in length
 gap of unknown length
 8798 9647: contig of 850 bp in length
 gap of unknown length
 9648 10555: contig of 908 bp in length
 gap of unknown length
 10556 11387: contig of 832 bp in length
 gap of unknown length
 11388 12285: contig of 898 bp in length
 gap of unknown length
 12286 13144: contig of 859 bp in length
 gap of unknown length
 13145 14010: contig of 866 bp in length
 gap of unknown length
 14011 14869: contig of 855 bp in length
 gap of unknown length
 14870 15756: contig of 887 bp in length
 gap of unknown length
 15757 16611: contig of 855 bp in length
 gap of unknown length
 16612 17482: contig of 871 bp in length
 gap of unknown length
 17483 18330: contig of 848 bp in length
 gap of unknown length
 18331 19206: contig of 876 bp in length
 gap of unknown length
 19207 20072: contig of 866 bp in length
 gap of unknown length
 20073 20946: contig of 874 bp in length
 gap of unknown length
 20947 21791: contig of 845 bp in length
 gap of unknown length
 21792 23944: contig of 2153 bp in length
 gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT
 TITLE
 JOURNAL
 COMMENT
 Shaffer, T. and Hood, L.
 Sequencing of human chromosome 14q24.3 region
 2 (bases 1 to 168919)
 Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,
 James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,
 Shaffer, T. and Hood, L.
 Direct Submission
 Submitted (30-MAY-1999) Multimegabase Sequencing Center, University
 of Washington, PO Box 357730, Seattle, WA 98195, USA
 NOTE: his record contains 192 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.
 1
 965: contig of 965 bp in length
 gap of unknown length
 966 1866: contig of 901 bp in length
 gap of unknown length
 1867 2747: contig of 881 bp in length
 gap of unknown length
 2748 3608: contig of 861 bp in length
 gap of unknown length
 3609 4462: contig of 854 bp in length
 gap of unknown length
 4463 5345: contig of 883 bp in length
 gap of unknown length
 5346 6187: contig of 842 bp in length
 gap of unknown length
 6188 7072: contig of 885 bp in length
 gap of unknown length
 7073 7927: contig of 855 bp in length
 gap of unknown length
 7928 8797: contig of 870 bp in length
 gap of unknown length
 8798 9647: contig of 850 bp in length
 gap of unknown length
 9648 10555: contig of 908 bp in length
 gap of unknown length
 10556 11387: contig of 832 bp in length
 gap of unknown length
 11388 12285: contig of 898 bp in length
 gap of unknown length
 12286 13144: contig of 859 bp in length
 gap of unknown length
 13145 14010: contig of 866 bp in length
 gap of unknown length
 14011 14869: contig of 855 bp in length
 gap of unknown length
 14870 15756: contig of 887 bp in length
 gap of unknown length
 15757 16611: contig of 855 bp in length
 gap of unknown length
 16612 17482: contig of 871 bp in length
 gap of unknown length
 17483 18330: contig of 848 bp in length
 gap of unknown length
 18331 19206: contig of 876 bp in length
 gap of unknown length
 19207 20072: contig of 866 bp in length
 gap of unknown length
 20073 20946: contig of 874 bp in length
 gap of unknown length
 20947 21791: contig of 845 bp in length
 gap of unknown length
 21792 23944: contig of 2153 bp in length
 gap of unknown length

* 23945 24793: contig of 849 bp in length
* 24794 25663: contig of 870 bp in length
* 25664 26504: contig of 841 bp in length
* 26505 27381: contig of 877 bp in length
* 27382 28254: contig of 873 bp in length
* 28255 29126: contig of 872 bp in length
* 29127 29963: contig of 837 bp in length
* 29964 30822: contig of 859 bp in length
* 30823 31680: contig of 858 bp in length
* 31681 32538: contig of 858 bp in length
* 32539 33390: contig of 852 bp in length
* 33391 34255: contig of 865 bp in length
* 34256 35105: contig of 850 bp in length
* 35106 36030: contig of 925 bp in length
* 36031 36869: contig of 839 bp in length
* 36870 37734: contig of 865 bp in length
* 37735 38584: contig of 850 bp in length
* 38585 39448: contig of 864 bp in length
* 39449 40283: contig of 835 bp in length
* 40284 41169: contig of 886 bp in length
* 41170 42014: contig of 845 bp in length
* 42015 42927: contig of 913 bp in length
* 42928 43788: contig of 861 bp in length
* 43789 44650: contig of 862 bp in length
* 44651 45491: contig of 841 bp in length
* 45492 46383: contig of 892 bp in length
* 46384 47232: contig of 849 bp in length
* 47233 48096: contig of 864 bp in length
* 48097 48951: contig of 855 bp in length
* 48952 49831: contig of 880 bp in length
* 49832 50676: contig of 845 bp in length
* 50677 51534: contig of 858 bp in length
* 51535 52388: contig of 854 bp in length
* 52389 53266: contig of 878 bp in length
* 53267 54108: contig of 842 bp in length
* 54109 55136: contig of 1028 bp in length
* 55137 55986: contig of 850 bp in length

* 55987 56874: gap of unknown length
* 56875 57721: gap of unknown length
* 57722 58587: gap of unknown length
* 58588 59427: gap of unknown length
* 59428 60292: contig of 865 bp in length
* 60293 61149: gap of unknown length
* 61150 62022: gap of unknown length
* 62023 62858: gap of unknown length
* 62859 63733: gap of unknown length
* 63734 64579: gap of unknown length
* 64580 65451: gap of unknown length
* 65452 66289: gap of unknown length
* 66290 67174: gap of unknown length
* 67175 68020: gap of unknown length
* 68021 68903: gap of unknown length
* 68904 69757: gap of unknown length
* 69758 70628: gap of unknown length
* 70629 71466: gap of unknown length
* 71467 72361: gap of unknown length
* 72362 73206: gap of unknown length

Query Match 16.1%; Score 73.6; DB 35; Length 168919;
Best Local Similarity 49.5%; Pred. No. 0.0031;
Matches 190; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 53 GAGGCGGAGGCGGAGACCGGCTGCTGAGAGACTGGGGGGGAGAGTCCCGGAGACCTAC 112
DB 42534 GGGGGGGGGGGCGGG 42533
QY 113 CCAAGCGCGGGGAGTACCTCCCGCTCACACCTCGGGCTGACATTCCTGGTGGCCCGCA 172
DB 42594 CGGGGGCGGTGGCGGG 42593
QY 173 GACGCTGCGCCCGCTGAGAGATGCGGGGGGAGCGGGGGCGGGGGGGGGGGGGCTGTC 232
DB 42654 GGGGGGTGTCGG 42713
QY 233 ACGTGAGAGCGCGCGCGGG 292
DB 42714 GGGGGGGGGGGCGGG 42773
QY 293 CGCGAGCGGAGGGGAGGGAGGTGCTGTGGCCGAGCGGAGCGGCGGAGCGAGCGAG 352
DB 42774 GCGGGGGCGGG 42833
QY 353 ACGTGAGCGCGCGCGGTGCTGACAGCTCGGGAGCTCAACATGCGTCTCCCGGGAGG 412
DB 42834 GCGGGGGGGGAGGG 42893
QY 413 CGTGTGCTGCGGCTGCGCGGCTC 436
DB 42894 GGGCGGCGGG 42917

RESULT 5
CEY39B6/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

MENT

Direct Submission
Submitted (09-JUN-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or twenematode.wustl.edu
On Jun 11, 1999 this sequence version replaced gi:4725958.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
Location/Qualifiers
1..342379
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y3986"

BASE COUNT 106570 a 59268 c 58805 g 109715 t 8021 others
ORIGIN

Query Match 15.3%; Score 70; DB 34; Length 342379;
Best Local Similarity 50.9%; Pred. No. 0.0076;
Matches 166; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 43 GGCTGTGCGAGAGCGCGAGAGCGCGCTGCTGCTGAGTGGGGTGAAGTGGC 102
DB 330422 GG 330363
QY 103 CGAGAGCTACCCAGAGCGCGGAGTACCTCCGCTACACCTCGGCTGCACTTCCCTGGG 162
DB 330362 GG 330303
QY 163 TGGCCCGCGAGAGCGCGCGCGCGCGCTGAGAGTATGCGCGCGCGAGCGCGGGCG 222
DB 330302 AGGG 330243
QY 223 GGGGCTGCTGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 282
DB 330242 GG 330183
QY 283 CTTAAAGCGCGCGAGCGCGAGCGAGTCTGTGTGCGCGAGCGCGAGCGCGCGC 342
DB 330182 GGG 330123
QY 343 GACAGCGGAGAGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
DB 330122 GCCGG 330097

RESULT 6

HS2IEA
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES

Location/Qualifiers
1..1072
/organism="human herpesvirus 2"
/db_xref="taxon:10310"
prim_transcript 95..>1072
/note="IE-5 mRNA and Intron"
Intron 344..866
/note="IE-5 Intron"
CDS 897..>1072
/note="Immediate early protein 5"
/codon_start=1
/protein_id="AA45849.1"
/db_xref="PID:G330284"
/db_xref="GI:330284"
/translation="MSMALKTDMPSDSRCTHRYGVDCAEIHKREREDPAATAV
TDEPLPLDPPVVRP"

BASE COUNT 132 a 322 c 543 g 75 t
ORIGIN

Query Match 14.9%; Score 68; DB 17; Length 1072;
Best Local Similarity 50.6%; Pred. No. 0.098;
Matches 164; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 37 GCCCTGTGCGCTGCGAGAGCGCGAGAGCGCGCTGCTGCTGAGTGGGGTGA 96
DB 366 GCCGGGGGAGACGGCGCGGGGGGAGCGGGCGGGGAGCGGGCGGGCGGGGG 425
QY 97 GTGCGCGAGCGCTACCCAGAGCGCGGAGTACCTCCGCTACACCTCGGCTGCACTTC 156
DB 426 GACGGGCGCGGGGGAGCGGCGCGGGGGAGCGGGCGGGGGAGCGGGCGGGGG 485
QY 157 CTTGGGTGCGCGCGAGAGCTGCTGCTGAGAGTATGCGGGCGGGGAGCGGGGG 216
DB 486 GCCGGGGGCGGGGGGAGCGGGGGGAGCGGGCGGGGAGCGGGCGGGGGGGA 545
QY 217 GGGGCGGGGCTGCTGCTGAGAGCGCGCGGGGGCGGGGGCGGGGGCGCGCGCG 276
DB 546 CGGGGGGAGCGGGGGGAGCGGGGGGAGCGGGGGGCGGGGGGAGCGGGGGGG 605
QY 277 CGGCTCTTAAAGCGCGAGAGCGCGAGAGTCTGTGTGCGCGAGCGCGAGCGC 336
DB 606 GGGGAGCGGGCGGGGGGAGCGGGGGGAGCGGGGGGCGGGGGGAGCGGGGGGG 665
QY 337 CGGGGGAGAGCGGAGAGCTGAG 360
DB 666 CGGGGGGAGCGGGGGGAGCGGGGGG 689

RESULT 7

LOCUS
DEFINITION
HS2IE
1560 bp DNA
VRL
15-MAR-1990
Herpes simplex virus type 2 Immediate-early (IE4) protein mRNA, 5'

end.
 M29384
 NID 9330281
 VERSION M29384.1 GI:330281
 KEYWORDS
 SOURCE
 ORGANISM
 Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed in baby hamster kidney clone 21 (Cl3) cells.
 human herpesvirus 2
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Alphaherpesvirinae; Simplexvirus.
 1 (bases 1 to 1560)
 Whitton, J. L. and Clements, J. B.
 The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide-coding regions of two spliced immediate-early mRNAs
 J. Gen. Virol. 65, 451-466 (1984)
 84137573
 JOURNAL MEDLINE
 FEATURES
 source
 Location/Qualifiers
 1..1560
 /organism="human herpesvirus 2"
 /db_xref="taxon:10310"
 prim_transcript 95..>1560
 /note="IE-4 mRNA and intron"
 Intron 344..906
 /note="IE-4 intron"
 CDS 969..>1560
 /note="immediate early protein 4"
 /codon_start=1
 /protein_id="AA04584.1"
 /db_xref="GI:555152"
 /db_xref="PID:955152"
 translation="MADIPDPALDTPPANHAPSPPLGSRKRRPYLPSSSEEGK
 /PIESSESTSESEDEVDLGRARRPRELGRFGLDSESTGTGSEGTGSDDD
 DDDSDGWLDPMPKSKRPRLNLTSPDRAGVPEVWRSDRPIRAAQPAPAQ
 SSGDAHAPRRAQRQMRSAATLDTLHYRQCNQ"
 BASE COUNT 222 a 480 c 714 g 144 t
 ORIGIN
 Query Match 14.4% Score 65.8; DB 17; Length 1560;
 Best Local Similarity 51.3%; Pred No. 0.17;
 Matches 177; Conservative 0; Mismatches 167; Indels 1; Gaps 1;
 25 TGGCCCTTGGGACCCCTGCGCTGCGACAGAGCGCGACCGCGCTCGTGAG 84
 350 TGAACCGGCGCGCGCGCGGAGAGCGCGCGGAGAGCGCGCGCGCGCGG 409
 85 ACTGGGGGTGGAGGTCCCGGAGCGTACCGAGCGCGGAGTACTCCGCTCACACTC 144
 410 ACGGCGCGGAGGAGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 469
 145 GGGCTGCATTCCTCGTGGCGCGCGGAGCGCTGCGCGCGCGCGCGCGCGCG 204
 470 GGGGCGCGGCG 529
 205 GGGGAGCGGCG 264
 529 GGGGCGCGGCG 588
 265 GGGGCGCGCGCGCGCGCTCTTAAGCGCGCGCGCGCGCGCGCGCGCGCG 324
 589 GCG 648
 325 GCAAGCG 369
 649 GGAACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
 RESULT 8
 LOCUS HSGG1 43058 bp DNA PRI 19-MAR-1997
 DEFINITION Human DNA sequence from cosmid GGI from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains

alpha and zeta globin genes and ESTs.
 284721
 NID 91817575
 VERSION 284721.1 GI:1817575
 KEYWORDS 16p13.3; alpha-globin; globin; zeta-globin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 43058)
 Flint, J. and Higgs, D. R.
 Direct Submission
 Submitted (28-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, UK. E-mail enquiries: humpdesanger.ac.uk
 IMPORTANT: This sequence is the entire insert of clone GGI. This
 clone was sequenced at the Institute of Molecular Medicine. The
 true left end of clone GGI is at 1 in this sequence. The true right
 end of clone GGI is at 43058.
 GGI is from a 280kb clone contig extending from the telomere of
 16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology
 Unit, Institute of Molecular Medicine, Oxford.
 GGI came from the Los Alamos, flow sorted human Chromosome 16
 library.
 FEATURES
 source
 Location/Qualifiers
 1..43058
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p13.3"
 /clone="GGI"
 1..72
 /note="FRAM repeat: matches 98.166 of consensus"
 97..399
 /note="Alusg repeat: matches 1.300 of consensus"
 516..811
 /note="AlusC repeat: matches 1.294 of consensus"
 823..977
 /note="L1 repeat: matches 5163.5328 of consensus"
 904..1028
 /note="L1ME3A repeat: matches 2.146 of consensus"
 1109..1760
 /note="L1MB8 repeat: matches 134.862 of consensus"
 1763..2057
 /note="AluY repeat: matches 2.297 of consensus"
 2443..2607
 /note="AluY repeat: matches 297.133 of consensus;
 incomplete repeat"
 2608..2905
 /note="AluY repeat: matches 297.1 of consensus"
 4231..4328
 /note="MIR repeat: matches 182.85 of consensus"
 4691..4989
 /note="AluY repeat: matches 7.301 of consensus"
 5222..5528
 /note="Alusg repeat: matches 1.300 of consensus"
 5530..5698
 /note="Alusg repeat: matches 130.296 of consensus;
 incomplete repeat"
 5714..6184
 /note="L1MB5 repeat: matches 922.459 of consensus"
 6320..6503
 /note="L1MB7 repeat: matches 2.208 of consensus"
 6531..6825
 /note="Alusg repeat: matches 295.1 of consensus"
 6864..7002
 /note="MTRC repeat: matches 1.136 of consensus"
 7045..7351
 /note="AluY repeat: matches 1.301 of consensus"
 7447..7545
 /note="MTRD repeat: matches 294.394 of consensus"
 7545..7885
 /note="L1MB3 repeat: matches 229.569 of consensus"
 7979..8281


```

repeat_region /note="AlusX repeat: matches 1. .302 of consensus"
8296. .8397 /note="Aluy repeat: matches 1. .301 of consensus"
repeat_region 8635. .8924 /note="AluDb repeat: matches 12. .299 of consensus"
repeat_region 8928. .8997 /note="L1M53 repeat: matches 631. .700 of consensus"
repeat_region 9052. .9327 /note="Aluy repeat: matches 13. .294 of consensus"
repeat_region 10382. .10428 /note="AlusX/9 repeat: matches 245. .291 of consensus"
repeat_region 10652. .10713 /note="L1M8 repeat: matches 920. .859 of consensus"
repeat_region 10714. .10966 /note="AlusP repeat: matches 1. .255 of consensus"
repeat_region 11022. .11168 /note="AlusX repeat: matches 1. .145 of consensus"
repeat_region 11169. .11463 /note="AlusP repeat: matches 1. .296 of consensus"
repeat_region 11466. .11635 /note="AlusX repeat: matches 133. .302 of consensus"
repeat_region 11730. .12317 /note="L1 repeat: matches 4729. .5317 of consensus"
repeat_region 12321. .12616 /note="Aluy repeat: matches 1. .296 of consensus"
repeat_region 12633. .12761 /note="AlusP repeat: matches 1. .131 of consensus"
repeat_region 12779. .13073 /note="AlusG repeat: matches 1. .297 of consensus"
repeat_region 13923. .14090 /note="AlusG repeat: matches 1. .297 of consensus"
repeat_region 14618. .15730 /note="Putative Cpg Island"
repeat_region 14964. .15071 /note="3 copies of 36 mer 83 & conserved"
repeat_region 15747. .15875 /note="MIR2 repeat: matches 145. .10 of consensus"
repeat_region 16241. .16541 /note="Aluy repeat: matches 301. .1 of consensus"
repeat_region 16892. .17027 /note="Aluo repeat: matches 1. .136 of consensus"
repeat_region 17029. .17326 /note="Aluy repeat: matches 1. .297 of consensus"
repeat_region 17368. .17506 /note="Aluo repeat: matches 10. .151 of consensus"
repeat_region 17945. .18086 /note="Aluo repeat: matches 153. .294 of consensus"
repeat_region 18106. .18426 /note="Aluo repeat: matches 1. .301 of consensus"
repeat_region 18715. .18833 /note="L1M87 repeat: matches 744. .865 of consensus"
repeat_region 18834. .19130 /note="AlusX repeat: matches 1. .296 of consensus"
repeat_region 19186. .19449

```

```

repeat_region /note="MER42c repeat: matches 1506. .1230 of consensus"
19431. .19593 /note="HVMARI repeat: matches 1221. .1058 of consensus"
repeat_region 19501. .19838 /note="AlusG repeat: matches 300. .63 of consensus"
repeat_region 19840. .19983 /note="AlusG repeat: matches 154. .300 of consensus"
repeat_region 20709. .20856 /note="4 copies of 37 mer 93 & conserved"
repeat_region 20881. .20952 /note="2 copies of 36 mer 99 & conserved"
repeat_region 20919. .21029 /note="3 copies of 37 mer 97 & conserved"
repeat_region 20996. .21067 /note="2 copies of 36 mer 99 & conserved"
repeat_region 21039. .21186 /note="4 copies of 37 mer 95 & conserved"
repeat_region 21216. .21323 /note="3 copies of 36 mer 97 & conserved"
repeat_region 21325. .21396 /note="2 copies of 36 mer 97 & conserved"
repeat_region 21360. .21433 /note="2 copies of 37 mer 97 & conserved"
repeat_region 21398. .21469 /note="2 copies of 36 mer 97 & conserved"
repeat_region 21471. .21542 /note="2 copies of 36 mer 99 & conserved"
repeat_region 21546. .21617 /note="2 copies of 36 mer 96 & conserved"
repeat_region 21579. .21726 /note="4 copies of 37 mer 94 & conserved"
repeat_region 22591. .22661 /note="AlusX repeat: matches 302. .232 of consensus"
repeat_region 22766. .23070 /note="AlusX repeat: matches 1. .302 of consensus"
repeat_region 23985. .24079,25345, .25549,25891, .26019 /note="start=1"
repeat_region 24135. .24680 /note="zeta-globin 2"
repeat_region 24135. .24680 /note="PID: e300371"
repeat_region 25208. .27584 /note="39 copies of 14 mer 90 & conserved"
repeat_region 25208. .27584 /note="Putative Cpg Island"
repeat_region 25208. .27584 /note="match: multiple ESTs; similar to HEMOGLOBIN"
repeat_region 25208. .27584 /note="match: 5' EST H57023 clone 204625; match: 5' EST"
repeat_region 25208. .27584 /note="match: 3' EST H67154 clone 211156"
Query Match 13.8% Score 63.2; DB 10; Length 43058;
Best Local Similarity 52.7%; Pred. No. 0.13;
Matches 137; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 24 CCGGCTTGAGGAGCCCTGAGCTGAGAGCGGAGAGCCGCTGAGTGA 83
DB 25589 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25648
QY 84 GACTGGGGGTGAGTCCCGGAGCGTACCGAGCGCGGAGTACCTCCGCTACACT 143
DB 25649 GCGCGGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25708
QY 144 CCGGCGGAGCTTCCTGAGTGGCGGCGGCGGAGAGCTGCGGCGGAGTGGCGG 203
DB 25709 GCGGCGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25768
QY 204 CCGGAGCGGCGGCGGCGGCGGCGGCTGCTACGTGAGAGAGCGGCGGCGGCGG 263
DB 25769 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25828
QY 264 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 283

```


Db 25829 GGGGTCGGGGCGGGGCC 25848

||||| ||||| ||||| ||||| |||||

RESULT 9

HUMBA3 2685 bp DNA PRI 18-SEP-1997

LOCUS Human alpha globin gene cluster on chromosome 16: psi2eta

DEFINITION pseudogene.

ACCESSION J00184

NID 9183792

VERSION J00184.1 GI:183792

KEYWORDS gene duplication; pseudogene; repeat region; zeta-globin.

SEGMENT 3 of 4

SOURCE Homo sapiens DNA.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Proudfoot, N.J., Gill, A., and Maniatis, T.

TITLE The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene

JOURNAL Cell 31 (3 Pt 2), 553-563 (1982)

MEDLINE 83129370

COMMENT The human alpha globin gene cluster, located on the short arm of chromosome 16, spans about 30 kb and includes the following five loci:

5'- zeta - pseudozeta - pseudalpha-1 - alpha-2 - alpha-1 -3'. The gene sequence shown below is considered a pseudogene because of the terminating codon at base 429. There are only two additional single-base differences between the coding sequences of the zeta and pseudozeta genes, although there are conspicuous differences in the 5' and 3' flanking regions. Moreover the first intron of this locus demonstrates length variation, apparently stemming from the repeat sequence 'acagctgggagggg' which is also found in the zeta gene and, in a similar form, in the 5' flank of the human insulin gene. The repeat sequence 'cgggg' is characteristic of the second introns of the zeta genes [1].

Typical promoter elements 'ccaat', and 'tata' are found at bases 291 and 326.

FEATURES

source 1. .2685

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="16p13.3"

356. .505

exon

/gene="phbz"

/number=1

join(411. .505,1767. .1971,2313. .2441)

CDS

/gene="phbz"

mutaton

429

/gene="phbz"

/note="c in phbz; g in hbz (see <humba3>)"

506. .1766

intron

/gene="phbz"

/number=1

565. .578

repeat_region

/note="repeat sequence copy 1"

1767. .1971

exon

/gene="phbz"

/number=2

1972. .2312

intron

/gene="phbz"

/number=2

1977. .1981

repeat_region

/note="repeat sequence, copy 1"

2313. .2541

exon

/gene="phbz"

/number=3

polyA_signal .2521. .2526

BASE COUNT 454 a 720 c 1155 g 356 t

ORIGIN About 1 kb after segment 2.

Query Match 13.8%; Score 63.2; DB 10; Length 2685;

Best Local Similarity 52.7%; Pred. No. 0.33;

Matches 137; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 24 CTGGCTTGGGCAACCCCTTGCCAGAGGCGGAGGCGGAGCCGCTCGTGA 83

Db 2011 CGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 2070

QY 84 GACTGGGGGTGAGATCCCGGAGCCTACCCAGCGCGGGAGTACCTCCGCTCACACT 143

Db 2071 GGGCGGGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2130

QY 144 CGGCTCAGTTCCTGCTGTGCGCGGAGAGCGTGGCGGCTGAGGATGCGGGG 203

Db 2131 GGGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 2190

QY 204 CGGGGAGCGGGCGGGCGGGCGGGCTGCTACGAGAGAGCGCGGGCGGGCGGGG 263

Db 2191 CGGGGCGGGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2250

QY 264 GGGGGCGGGCGGGCGGGCTCC 283

Db 2251 GGGGTCGGGGCGGGGCC 2270

RESULT 10

HSV2HG52

LOCUS HSV2HG52 154746 bp DNA VRL 04-DEC-1998

DEFINITION Herpes simplex virus type 2 (strain HG52), complete genome.

ACCESSION 2860099

NID 91869820

VERSION 286099.1 GI:1869820

KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyribonuclease B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; RL1 gene; RL2 gene; RL3 gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL14 gene; UL15 gene; UL16 gene; UL17 gene; UL18 gene; UL19 gene; UL2 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL3 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL4 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45 gene; UL46 gene; UL47 gene; UL48 gene; UL49 gene; UL49A gene; UL5 gene; UL50 gene; UL51 gene; UL52 gene; UL53 gene; UL54 gene; UL55 gene; UL56 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; uracil-DNA glycosylase; US1 gene; US10 gene; US11 gene; US12 gene; US2 gene; US3 gene; US4 gene; US5 gene; US6 gene; US7 gene; US8 gene; US8A gene; US9 gene

SOURCE human herpesvirus 2.

ORGANISM human herpesvirus 2.

REFERENCE 1. (bases 1 to 154746)

AUTHORS McGeoch, D.J., Moss, H.W., McNab, D., and Frame, M.C.

TITLE DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons

JOURNAL J. Gen. Virol. 68 (Pt 1), 19-38 (1987)

MEDLINE 87111457


```

REFERENCE          2 (bases 1 to 154746)
AUTHORS            Everett,R.D. and Fenwick,M.L.
TITLE              Comparative DNA sequence analysis of the host shutoff genes of
                  different strains of herpes simplex virus: type 2 strain HG52
                  encodes a truncated UL41 product.
JOURNAL            J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)
MEDLINE            90278430
REFERENCE          3 (bases 1 to 154746)
AUTHORS            McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
TITLE              Comparative sequence analysis of the long repeat regions and
                  adjoining parts of the long unique regions in the genomes of herpes
                  simplex viruses types 1 and 2
JOURNAL            J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
MEDLINE            92113549
REFERENCE          4 (bases 1 to 154746)
AUTHORS            Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
TITLE              A novel herpes simplex virus gene (UL49a) encodes a putative
                  membrane protein with counterparts in other herpesviruses
JOURNAL            J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
MEDLINE            92356101
REFERENCE          5 (bases 1 to 154746)
AUTHORS            Dolan,A.
TITLE              Direct Submission
JOURNAL            Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
                  Street, Glasgow, G11 5JR, UK
FEATURES
SOURCE             1. 134746
                  /organism="human herpesvirus 2"
                  /strain="HG52"
                  /db_xref="taxon:10310"
misc_feature       1 254
                  /note="a sequence"
misc_feature       235
                  /note="start of TRL (terminal copy of Long Repeat region)"
                  440..934
                  /gene="RL1"
                  /number=1
gene              440..126810
                  join(440..934,1089..1379)
                  /gene="RL1"
                  /codon_start=1
                  /protein_id="CAB06759.1"
                  /db_xref="PID:e308397"
                  /db_xref="PID:g1869821"
                  /db_xref="GI:1869821"
                  /db_xref="SWISS-PROT:P28283"
                  /translation="MSRRRGPRRRRPPGAPVPRGAPVPRPGALPTADSQM
                  VPAVDSGTAVESAPPAASSLKRMLVPAQADSDADADAGDDLEMANSPSPSEGGKAP
                  EAPHAAPAAACPPPPPKRERGPQRPPLPHALRLRTTEYLARLSRRRPPSPAD
                  APGKACFSPRVQVRLVAMETAAARLARSGSMARERDRFRFRVAAAVAGPCIE
                  PEARARARARAHEDGPAEEBAAAAGSSAAAGPGRRAV"
                  446..460
repeat_region     /rpt_family="7"
                  461..547
repeat_region     /rpt_family="6"
                  548..1070
repeat_region     /rpt_family="5"
                  1089..1379
exon              /gene="RL1"
                  /number=2
                  /product="neurovirulence factor"
                  1738..1743
polyA_signal      /gene="RL1"
                  2303..124947
gene              /gene="RL2"
                  2303..2377
exon              /gene="RL2"
                  /number=1
                  join(2303..2377,2785..3462,3644..5368)
                  /gene="RL2"
                  /codon_start=1
                  /protein_id="CAB06760.1"

                  /db_xref="PID:e304156"
                  /db_xref="PID:g1869822"
                  /db_xref="GI:1869822"
                  /db_xref="SWISS-PROT:P28284"
                  /translation="MEPRPGTSSRADGPPRRPPTGCTQPPAAPHANGMLNDQWLAS
                  SDSEETFEVIGISDDDLHRDSTSENGSIDTEFEGLMDAATPPAPRAERQSGTPPAD
                  AQSGGGGPGVEEAEAGGGDVCACVCTDITAPLRQSPFCLHPICTIPMKTPIPR
                  NTCPICNTPAAYLIVGTASGSEFTIIVDPRTVRVAEAVRAGTAVDIWTKNPT
                  APRSLSGHTVRLSPTPMPGDDDDDLADVDYPPAPRRARRGGGAGATRGCT
                  SOPATPAPPGAPRRSSSGGAPLRAVGSGSGGAPVAAVPRVASTPPAGGRQ
                  ARVGENAAAGRTPPAPRQAQEPPIYISDPPSPRRPAGCTLSVSSSAOY
                  SGGGCGGCTQSSGRRAPRAVAFAVRSPPRRAAIVVASADAAPAPAPVVDH
                  RAPSRTQQTDTQAQSLGRAGATDARGSGGPGPAEGPGVPRGTNPGAPAAEBA
                  AAREKRRGSDGPPAASSASASSAAPRSPAPGVGAKRAAPRPPDSDGRGHL
                  APAGAAPPSPSSOAAVAAASSASSASSASSASSASSASSASSASSASSASSA
                  SSSAGCGGGSVASSAGGEREMSLGPRAAPGPRCAKTRAEGBPPGARDPAP
                  GTRRYLEIAVSSVYALAPYVNTKYTDCLPVIDMETGIGATVVLVDQGNADLIR
                  AAPAMSRRLLEPHANRCVPPDPYPPPASEWNSLWTVGNMLFDQGLTVGLDPRH
                  GLRSRHPWSREOGAPAPADGAPAGHG"
                  2785..3462
exon              /gene="RL2"
                  /number=2
                  3468..3607
repeat_region     /rpt_family="4"
                  3644..5368
exon              /gene="RL2"
                  /number=3
                  /function="modulator of cell state and gene expression"
                  4655..4771
repeat_region     /rpt_family="3"
                  complement(4811)
misc_feature       /gene="RL2"
                  /note="proposed LAT splice acceptor site"
                  5618..5623
polyA_signal      /gene="RL2"
                  complement(7025)
misc_feature       /gene="RL2"
                  /note="proposed LAT splice donor site (5' end of stable
                  LAT species)"
                  complement(7732)
misc_feature       /gene="RL2"
                  /note="primary transcription initiation site for latency
                  associated transcripts (LAT's)"
                  9047..9193
repeat_region     /rpt_family="2"
                  9276..9297
repeat_region     /rpt_family="1"
                  9298
misc_feature       /gene="RL2"
                  /note="start of UL (Long Unique region)"
                  9427..10101
gene              /gene="UL1"
                  9427..10101
CDS               /gene="UL1"
                  /codon_start=1
                  /product="viralion glycoprotein L"
                  /protein_id="CAB06761.1"
                  /db_xref="PID:e304262"
                  /db_xref="PID:g1869823"
                  /db_xref="GI:1869823"
                  /db_xref="SWISS-PROT:P28278"
                  /translation="MGFVCLFGLVWGWAGWGAGGSGQATPEYVLRSVAKREVDILRVC
                  MRPAPDDVSMRYEAPVIDYKARIDGFLFLRPGGLDTPFLMDHAORATVYNFLRAG
                  FLEDSHVSVPADTQSTTTRKALVKEIRDALSKRQAOVSHAVVRAGVCVDFSLRRRC
                  VGRDLRPAINTSTWEPVSSDDEASSOSKPLATOPVYALSNAPRRVSPYGRRRH
                  TRLRRN"
                  10211..10978
gene              /gene="UL2"
                  10211..10978
CDS               /gene="UL2"
                  /codon_start=1
                  /product="uracil-DNA glycosylase"
                  /protein_id="CAB06762.1"

```



```

/db_xref="PID:e304157"
/db_xref="PID:g1869824"
/db_xref="GI:1869824"
/db_xref="SWISS-PROT:P28275"
/translation="MESASTPEQPLGSGDAPPLPTSPVLDWAERFALIDNMR
PLIPEPLANTARILAEVDRSCOTEFREPREVFEEDVAVYIIGDDPH
HFGAHGLAEFVRADYVPSPSLRVLAANKCYDAMSGCEKARAGVILNLT
LTVRGAASHSKLSMDPEVGVQRLAARGLVEMIMGAHNAONLRPDRPHYTK
FSHSPSLSKVPEFGCOHFLANRYLETREDIMPMDSV"
polyA_signal
11019..11024
/gene="RL2"
11033..11734
/gene="UL3"
11033..11734
CDS
11033..11734

Query Match      13.8%; Score 63; DB 17; Length 154746;
Best Local Similarity 48.7%; Pred. No. 0.092;
Matches 171; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

OY 37 GCCCTGCGCTGCGCAGAGCGCGGAGCCGAGAGCCCTCGCTGAGACTGGGGCTGA 96
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133270 GCCGGGGGAGCGCGCGCGGAGCGCGGCGCGGCGCGCGCGCGCGCGGCG 133329
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 97 GGTGCCCGAGCGGTACCCAGCCCGCGGAGTACTCCCGCTCAGCTCGGCGTGCAGTTC 156
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133330 GACGGGGCGCGGAGCGCGCGCGGAGCGCGGCGCGGAGCGCGGCGCGGCGG 133389
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 157 CTTGGGTGCGCCGCGAGAGCTGCGCCGCGGTGAGGATGCGCGGCGCGGAGCGGGGCG 216
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133390 GCCGGGGGCGCGCGCGCGGAGCGCGGCGCGGAGCGCGGCGCGGAGCGGGGAGAG 133449
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 217 GGGGGCGGGGCTCTGTCAGTGAAGAGCGCGCGGCGCGGCGCGGCGCGCGCGCC 276
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133450 GGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 133509
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 277 CGGCTCTTAAGCGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCG 336
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133510 GACGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGG 133559
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 337 CCGGGGAGACCGGAGACGCTGAGCGCGCGGCTCGCTCAGCTCGGGAGC 387
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 133570 CCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 133620
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 11
-----V2HG52/c
IS HSV2HG52 154746 bp DNA VRL 04-DEC-1998
INITIATION Herpes simplex virus type 2 (strain HG52), complete genome.
ACCESSION Z86099
VERSION Z86099.1 GI:1869820
KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyuridine triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; RLI gene; RL2 gene; RSL gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL14 gene; UL15 gene; UL17 gene; UL18 gene; UL19 gene; UL2 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL3 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL4 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45 gene; UL46 gene; UL47 gene; UL48 gene; UL49 gene; UL49a gene; UL5 gene; UL50 gene; UL51 gene; UL52 gene; UL53 gene; UL54 gene; UL55 gene; UL56 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; uracil-DNA glycosylase; US1 gene; US10 gene; US11 gene; US12 gene; US2 gene; US3 gene; US4 gene; US5 gene; US6 gene; US7 gene; US8 gene; US8a

```

```

SOURCE human; US9 gene.
ORGANISM human herpesvirus 2.
REFERENCE 1. (bases 1 to 154746)
AUTHORS McGeoch,D.J., Moss,H.W., McNab,D. and Frame,M.C.
TITLE DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons
J. Gen. Virol. 68 (Pt 1), 19-38 (1987)
JOURNAL MEDLINE 87111457
REFERENCE 2 (bases 1 to 154746)
AUTHORS Everett,R.D. and Fenwick,M.L.
TITLE Comparative DNA sequence analysis of the host shutoff genes of different strains of herpes simplex virus: type 2 strain HG52 encodes a truncated UL41 product
J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)
JOURNAL MEDLINE 90278430
REFERENCE 3 (bases 1 to 154746)
AUTHORS McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
TITLE Comparative sequence analysis of the long repeat regions and adjoining parts of the long unique regions in the genomes of herpes simplex viruses types 1 and 2
J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
JOURNAL MEDLINE 92113549
REFERENCE 4 (bases 1 to 154746)
AUTHORS Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
TITLE A novel herpes simplex virus gene (UL49A) encodes a putative membrane protein with counterparts in other herpesviruses
J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
JOURNAL MEDLINE 92356101
REFERENCE 5 (bases 1 to 154746)
AUTHORS Dolan,A.
TITLE Direct Submission
JOURNAL TITLE Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church Street, Glasgow, G11 5UR, UK
FEATURES
source
misc_feature
1..254
/note="a sequence"
misc_feature
255
/note="start of RLI (terminal copy of Long Repeat region)"
exon
440..934
/gene="RL1"
number=1
440..126810
/gene="RL1"
join(440..934,1089..1379)
/gene="RL1"
/codon_start=1
/protein_id="CAP06759.1"
/db_xref="PID:e308397"
/db_xref="PID:g1869821"
/db_xref="PID:g1869821"
/db_xref="SWISS-PROT:P28283"
/translation="MSRRGPRRGRPRRPGAPVPRPGAPVPRPGALPTADSQM
VPADSGAVSAPASASILRRWLLVPGADSDADAGNDAAEANSPPSGGKAP
EAPHAAPAAAPPPPRKRRKRRPRLPHLLRLRTTTEYLARLSRRRRPASPAD
APRGKVCSPVQVRLVAVETAKRLARSGSWAKERKDRFRVRVAALAEVIGPCLE
PEARARVCAARAHEDGGPAEEAARSSAAAGPGRRAV"
repeat_region
446..460
/rpt_family="7"
repeat_region
461..547
/rpt_family="6"
repeat_region
943..1070
/rpt_family="5"
exon
1089..1379
/gene="RL1"

```


polyA_signal /product="neurotrophin factor"
1738..1743
/gene="RL1"
gene 2303..124947
/gene="RL2"
exon 2303..2377
/gene="RL2"
/number=1
join(2303..2377,2785..3462,3644..5368)
/gene="RL2"
/codon_start=1
/protein_id="CAB06760.1"
/db_xref="PID:e304156"
/db_xref="PID:g1869822"
/db_xref="GI:1869822"
/translation="MEPRPTGSRADPPEPRPPTGCTGQAPAHAMNTLNDQMTLAS
/SSESETEVIGISDDLHRDSTSEAGSTETEMFPAGLMDAAPAPPRPERGSGSTPAD
AOGSCGGGVGEENAEAGGGGVCAVCTDEIAPLRCSFCLPFCTPCMKTTIPLR
NCPCLNCPVVALIVGVASGSFSTIPVNDPRTVEEAARVAGTAVDEFTMGNTPT
APRSLSIGHTVTRALSPTPMPMGCTDEDDELDADVYPAPRRAPRRGGAGATGRT
SOPATRRAPPCAPRRSSSGGAPLRAGVSGSGGGPVAAYVPVVASLPAGAGGRAD
ARVEDAPAAAGRTPPARAPPAOEPRTIYSDSPSPRRAPAGPLSTSVSSAOV
SSGPGGGLPOSSGGAAPRAAPVAPRVASPPRAAPAVYVSAADAGAPVAVDAAH
RAPRRKMTQADTDQASLGRAGATDANGSGPGEGGPPVKTNPGAAPHAEEA
AARPKRRSGDSGPASASSASAPRSPPLAPQCGAKRAAPRAPDSDGRHGEPL
APASAGAPSPSPSSQAAVVAASSSSASSSSSADVYAPRRAPRRGGAGATGRT
SSAGAGGSPVSAAGGERETSLGPRAAAPRGCARKTTRAEGGPPGAPAPAP
GLTRPLIAGVSSVVALAPYKNTVTGCLVLYDMETGHIGATYVLYDQGNVADLR
AAAPWMSRTLLPEHARNCVRPDIPTPPASEMNSLMTWTPVGNMLFDQGLTVGALDEPH
GIRSRHPSREGGAPAPADAPAGHE"
exon 2785..3462
/gene="RL2"
/number=2
repeat_region 3468..3607
/rpt_family="4"
exon 3644..5368
/gene="RL2"
/number=3
/function="modulator of cell state and gene expression"
repeat_region 4655..4771
/rpt_family="3"
misc_feature complement(4811)
/gene="RL2"
/note="proposed LAT splice acceptor site"
polyA_signal 5618..5623
/gene="RL2"
misc_feature complement(7025)
/gene="RL2"
/note="proposed LAT splice donor site (5' end of stable
LAT species)"
misc_feature complement(7722)
/gene="RL2"
/note="primary transcription initiation site for latency
associated transcripts (LAT's)"
repeat_region 9047..9193
/rpt_family="2"
repeat_region 9276..9297
/rpt_family="1"
misc_feature 9298
/gene="RL2"
/note="start of UL (long unique region)"
gene 9427..10101
/gene="UL1"
9427..10101
/codon_start=1
/product="villon glycoprotein L"
/protein_id="CAB06761.1"
/db_xref="PID:e304263"
/db_xref="PID:g1869823"
/db_xref="GI:1869823"

[illegible]

JOURNAL
REFERENCE
TITLE
AUTHORS
JOURNAL
JOURNAL
Submitted (13-NOV-1995) Sherry Leonard, University of Colorado
Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave,
Denver, CO 80262, USA

FEATURES
source
Location/Qualifiers
1. .1977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pBSshp1a7"
/clone_id="Clontech lambda gt10 cDNA library: human
hippocampus"
/sex="male"
/tissue_type="hippocampus"
/dev_stage="20 yr old adult"
1. .7
8. .73
8. .1516
/codon_start=1
/product="alpha 7 neuronal nicotinic acetylcholine
receptor"
/protein_id="AA83561.1"
/db_xref="PID:g1125077"
/db_xref="GI:1125077"
/translation="MRCSPGVWLAALSLHVSLOGEGRKLYKLYKNYPLERPV
ANDSOPLIVFSLNLOIMVDKRNQVLTINILQMSWIDHYLKNYSEPKYVRE
PDQIMKPDILLYNSADERDQTEHFNVLNPGSCQYLPFGKSSYIDVPEPD
VOHCKLFGSGSYGMSIDLOMOEDISGYPNGEMDLYIPKREBERECCKEEP
DNYETVMRRRTLYGYNLIRPVILSALALYFLPADSGEISGITYLSITTFM
LYAEIPLRTSDSPLTAQYFASITMIVGLSVVYVYALDTHHDPDGGKRMKVI
LWMCWFLMKRPRGKEDKPRACQHQKRSLSASVEASVAPPASNGMLTYGFGVL
DGVHCVPTDPSGVYCGRMASPTNDEHILHGQPPEDDPLARILEVERIANRRCQ
DESEAVCEWKFACVVDRLCLMAFVFTICTIGILMSAPNFEVAVSKDEA"
1514. .1977
1954

3'UTR
polya_site
BASE COUNT 426 a 567 c 524 g 460 t
ORIGIN

Query Match 13.6% Score 62; DB 10; Length 1977;
Best Local Similarity 100.0%; Pred. NO. 0.54;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

386 ACTCAACATGCGCTGCTGCCGCGGAGCGCTGCGCTGCGCGCTGCTGCA 445
|||||
Db 1 ACTCAACATGCGCTGCTGCCGCGGAGCGCTGCGCTGCGCGCTGCTGCA 60
|||||

QY 446 CG 447
||
Db 61 CG 62

RESULT 13
LOCUS SCE94 38532 bp DNA BCT 12-APR-1999
DEFINITION Streptomyces coelicolor cosmid E94.
ACCESSION AL049628
NID 94585581
VERSION AL049628.1 GI:4585581
KEYWORDS acetyltransferase; ATP-binding; Clp-family ATP-binding protease;
cseB; cseC; DNA repair; DNA-binding; ECF sigma factor; exporter;
L-aspartate oxidase; lipA; lsr2; mlt; nadB; nadC;
nicotinate-nucleotide pyrophosphorylase; pantoate-amino acid ligase;
putative adenine glycosylase; response regulator; sensory histidine
kinase; sigE.
Streptomyces coelicolor.
Streptomyces coelicolor.
Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 38532)

AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
JOURNAL
Submitted (12-APR-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
CDS are numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4781(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nhn.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optionally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid E94 lies between E7 and E126 on the Aser-E genomic
restriction fragment.

FEATURES
source
Location/Qualifiers
1. .38532
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid E94"
complement(1..1512)
/gene="SCE94.01c"
1. .105
/gene="SCE94.01c"
/note="Nominal overlap with Streptomyces coelicolor E7."
complement(1..1512)
/gene="SCE94.01c"
/note="SCE94.01c, partial CDS unknown, alanine-rich
protein, len: >503aa"
/codon_start=1
/transl_table=11
/label="SCE94.01c
/product="alanine-rich hypothetical protein"
/protein_id="CAB40850.1"
/db_xref="PID:e1424928"
/db_xref="PID:94585582"
/db_xref="GI:4585582"

CH19R30879/c CH19R30879 41936 bp DNA PRI 01-APR-1997
 LOCUS Homo sapiens DNA from chromosome 19-cosmid R30879 containing USF2,
 DEFINITION genomic sequence.
 ACCESSION AD000684
 NID 91905917
 VERSION AD000684.1 GI:1905917
 KEYWORDS chromosome 19; transcription factor.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 41936)
 Lamerdin,J.E.
 DIRECT SUBMISSION
 JOURNAL Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center,
 Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
 CA, USA, 94551 janelacgt.llnl.gov owetornak.llnl.gov
 GSUB:S:1010600.
 COMMENT Human Genome Center
 Biology and Biotechnology Research Program
 Lawrence Livermore National Laboratory
 7000 East Avenue
 Livermore, CA 94550 USA
 constructed at LLNL from flow-sorted chromosomes
 from hybrid 5HL2-B, which carries chromosome 19 as its only human
 chromosome.

FEATURES
 source 1. 41936
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /cell_line="5HL2-B"
 /cell_type="fibroblast"
 /clone="R30879"
 /clone_lib="LL19NC03 R chromosome 19-specific cosmid
 library"
 complement(1..34)
 /note="repeat match = HSA115306; putative"
 /rpt_family="Alu"
 complement(811..1099)
 /note="repeat match = HSA100208; putative"
 /rpt_family="Alu"
 1335..1479
 /note="repeat match = HSA105512; putative"
 /rpt_family="Alu"
 1367..2305
 /rpt_family="Alu"
 /note="repeat match = HSA102043; putative"
 /rpt_family="Alu"
 complement(3093..3356)
 /note="repeat match = HSA102254; putative"
 /rpt_family="Alu"
 4030..4168
 /note="repeat match = HSA103448; putative"
 /rpt_family="Alu"
 /note="repeat match = HSA10449; putative"
 4031..4288
 /note="repeat match = HSA101019; putative"
 /rpt_family="Alu"
 4435..4533
 /rpt_family="Alu"
 /note="repeat match = HSA102522; putative"
 /rpt_family="Alu"
 4448..4565
 /note="repeat match = HSA101715; putative"
 /rpt_family="Alu"
 4926..5218
 /note="repeat match = HSA106523; putative"
 /rpt_family="Alu"
 complement(5804..6091)
 /note="repeat match = HSA104878; putative"
 /rpt_family="Alu"
 complement(6051..6102)
 /note="repeat match = HSA104290; putative"
 /rpt_family="Alu"

repeat_region 7131..7434
 /note="repeat match = HSA102525; putative"
 /rpt_family="Alu"
 repeat_region 7137..7589
 /note="repeat match = HSA100829; putative"
 /rpt_family="Alu"
 7911..16824
 /gene="L1SCH7"
 /gene="L1SCH7"
 151n(<7911..8030,10891..10947,11512..11658,15325..15498,
 15802..15933,16075..16700,16790..16824)
 /function="human homolog of Mus musculus B6CBA L1sch7
 mRNA"
 /note="putative"
 /codon_start=2
 /product="liver-specific bHLH-Zip transcription factor"
 /protein_id="AB51178.1"
 /db_xref="pid:91905918"
 /db_xref="GI:1905918"
 /translation="ADLTEDQATMGDSGVYGSVSNADLQGNNAEVALIVLGRSG
 VALLPFGQAPTEGWLFTVYVCLAFILITLLIGMCCCPHTCCCVCCPCPCPC
 CCEPARYAAGKAATSGVPSIYAPSTIAHLSPATPPPPATPPGAPANGPGGYPGV
 DRSSGVSQGYRIQASQODSMKLYLMEKELANFDESRPGSPGSRGVRKMSYTSB
 EDMRSRPSRGPALTPIDEMGSHSPSRPGMDPEAPRQAGGWARPRKASVDA
 LDLTPTAESGSRSPISNGRSRAVMPRPSRDLDYQDDSDRPPRRDPYDDE
 RSRPRPADRPSHHHRRDRDRDNGSRGDLPYDGRLLLEAVRKKGSEERRRPHKEEE
 EAYPPAPPPYSETDSQASRERLKNLALRESLYV"
 8303..8386
 /note="repeat match = HSA105819; putative"
 /rpt_family="Alu"
 8325..8609
 /note="repeat match = ALU; putative"
 /rpt_family="Alu"
 complement(9257..9546)
 /note="repeat match = HSA105084; putative"
 /rpt_family="Alu"
 complement(9513..9555)
 /note="repeat match = HSA100054; putative"
 /rpt_family="Alu"
 9599..9759
 /note="repeat match = HSA117230; putative"
 /rpt_family="Alu"
 9639..9813
 /note="repeat match = HSA106130; putative"
 /rpt_family="Alu"
 10359..10498
 /note="repeat match = HSA106554; putative"
 /rpt_family="Alu"
 11510..11657
 /gene="L1SCH7"
 /gene="L1SCH7"
 /note="similarity: gb|H42128|H42128: y061b12.r1 Homo
 sapiens cDNA clone 18339 5'; putative"
 11957..12072
 /note="repeat match = HSA10157; putative"
 /rpt_family="Alu"
 11981..12253
 /note="repeat match = ALU; putative"
 /rpt_family="Alu"
 12168..12279
 /note="repeat match = HSA103934; putative"
 /rpt_family="Alu"
 complement(12334..12610)
 /note="repeat match = HSA100279; putative"
 /rpt_family="Alu"
 complement(12391..12982)
 /note="repeat match = HSA102441; putative"
 /rpt_family="Alu"
 complement(12645..13088)
 /note="repeat match = HSA102504; putative"
 /rpt_family="Alu"
 16521..16586
 /gene="L1SCH7"
 /note="similarity: gb|R39945|R39945: yf52e09.s1 Homo

misc_feature

[illegible]

Db	18009	CGAGGAGGAGGGAGAGGGGCGGCCGCCCCGAGCCGCGCTCACGCCGCGC	17950
Qy	295	CGAGCCGACGGCGAGGTGCCTCTGTGGCCGACGAGCGCCGCGGACCA	346
Db	17949	GCAGGAGGGGACGAGAGATACGGGAGCCGCGTCCGTGATCACGGGACA	17898
RESULT	15		
HUSUF2/c			
LOCUS	HUSUF2	14440 bp	DNA
DEFINITION	H.sapiens USF2 gene.		PRI
ACCESSION	X07661		26-JUN-1997
NID	G1806093		
VERSION	X07661.1	GI:1806093	
KEYWORDS	USF2 gene.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 14440) Groenen,P.M., Garcia,E., Debeer,P., Devriendt,K., Frys,J.P. and Van de Ven,W.J. Structure, sequence, and chromosome 19 localization of human USF2 and its rearrangement in a patient with multicystic renal dysplasia Genomics 38 (2), 141-148 (1996)		
TITLE	2 (bases 1 to 14440) Groenen,P. Direct Submission		
JOURNAL	Submitted (27-MUG-1996) P. Groenen, Center for Human Genetics, Laboratory for Molecular Oncology, Herestraat 49, 3000 Leuven, Belgium		
MEDLINE	Related sequence: X90823-X90826. Location/Qualifiers		
REFERENCE	1. 14440		
AUTHORS	/organism="Homo sapiens"		
JOURNAL	/db_xref="taxon:9606"		
TITLE	/chromosome="19"		
MEDLINE	/haplotype="2n"		
REFERENCE	/map="q13.1"		
AUTHORS	/clone_lib="LIND 7058"		
JOURNAL	/dev_stage="adult"		
MEDLINE	join(3688..3749,4031..4077,4166..4284,4388..4588,		
REFERENCE	5032..5182,5303..5390,5668..5726,13284..13378,		
AUTHORS	13532..13660,13753..14380)		
JOURNAL	/gene="USF2"		
TITLE	3688..14380		
MEDLINE	/gene="USF2"		
REFERENCE	3688..3749		
AUTHORS	/gene="USF2"		
JOURNAL	/number=1		
TITLE	join(3688..3749,4031..4077,4166..4284,4388..4588,		
MEDLINE	5032..5182,5303..5390,5668..5726,13284..13378,		
REFERENCE	13532..13660,13753..13842)		
AUTHORS	/gene="USF2"		
JOURNAL	/codon_start=1		
TITLE	/protein_id="CA68942.1"		
MEDLINE	/db_xref="PDB:e291919"		
REFERENCE	/db_xref="PID:g1806094"		
AUTHORS	/db_xref="GI:1806094"		
JOURNAL	/db_xref="SWISS-PROT:O15853"		
TITLE	/translation="MDM2DGLDPAASSTATAAASHDKPEAEKGVELDGSDPGAE		
MEDLINE	ETLAATISVQAALFGNDNITORTFTENGQVTYRVAVYVDQLDGDGTAAVSIV		
REFERENCE	STAAPRGQDAVTVGVNGAARGPPAAAYVPGPAPFLAVLQNFSGSGSPAAY		
AUTHORS	VSGEERAFVPPASSVGDTTAVSYQTQDSLOAGGFYMMTPDVLTGTGTIAPT		
JOURNAL	HPYSKIDGTRTPDERRAOHNEVERRRDKINNVIWOLSKTIIPDCNNASTGASK		
TITLE	GHLSKADYIRLELRONTNRQGEFKFAEHLQMDNELRROIIEKKENALLRLAQOQ		
MEDLINE	HNLEWVGEGTRO"		
REFERENCE	3750..4030		
AUTHORS	/gene="USF2"		
JOURNAL	/number=1		
TITLE	4031..4077		

us-08-956-518a-94.png

ALIGNMENTS

last;

(without alignments)
268.910 Million cell updates/sec

Sequence: 1 AGAACGCAAGGGAGAGGTAG.....CTCCTGCACGGTAAGCCAC 457

Searched: 311585 seqs, 125096042 residues

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
C	1	127	27.8	1876	1	T48239	Neutroal nictotinic	
C	2	127	27.8	1876	1	V12197	Human neuronal nict	
C	3	69.2	15.1	3861	1	Q51543	Lipopolyascharidase	
C	4	68.8	15.1	114955	1	X53491	Human adenosine A	
C	5	65.2	14.3	143955	1	X53491	Human adenosine A	
C	6	64	14.0	3198	1	X02874	Human IL-1ra BAC c	
C	7	60.6	13.3	1281	1	Q23296	Human IL-1ra BAC c	
C	8	57.8	12.6	11820	1	V18130	Human chromosome 1	
C	9	56.4	12.3	12001	1	Q76213	HSV L/ST region. H	
C	10	56	12.3	1130	1	V60339	cDNA sequence of f	
C	11	53	11.6	4257	1	V10362	Infected cell prot	
C	12	53	11.6	4257	1	V68520	The nucleotide seq	
C	13	52.4	11.5	12001	1	Q76213	HSV L/ST region. H	
C	14	52.4	11.5	5228	1	V81384	Nucleic acid encoc	
C	15	51.8	11.3	4403	1	Q47927	Paired basic amin	
C	16	51.8	11.3	2218	1	Q47929	Paired basic amin	
C	17	51	11.2	6225	1	X55273	Human enzyme-rela	
C	18	50.8	11.1	407	1	V44430	Human Mycobacteriu	
C	19	50.8	11.1	407	1	V64539	M. tuberculosis int	
C	20	50.4	11.0	7135	1	V50431	Streptomyces clavos	
C	21	50.2	11.0	1353	1	Q23295	HSV-1 (F) ICP34.5	
C	22	50	10.9	4067	1	Q63192	Human adeno-associ	
C	23	50	10.9	7011	1	V20464	Human L-myc oncoge	
C	24	49.6	10.9	1303	1	Q23297	HSV-1 (CVG-2) ICP	
C	25	49.2	10.8	701	1	Q76209	HSV L/ST ORF1. Her	
C	26	49.2	10.8	117213	1	V62176	HSV-2 strain SB5 C	
C	27	49	10.7	4405	1	Q25444	PACE composite seq	
C	28	49	10.7	4405	1	T05556	Human PACE coding	
C	29	49	10.7	4488	1	T35520	Human thrombopoiet	
C	30	49	10.7	223	1	T66412	Wild type SNF-5' u	
C	31	49	10.7	223	1	V01700	Human T-cell leuka	
C	32	48.4	10.6	801	1	V73801	Human T-cell leuka	
C	33	48	10.5	4020	1	T91361	Orf virus genomic	
C	34	48	10.5	117213	1	V62176	HSV-2 strain SB5 C	
C	35	47.8	10.5	1032	1	Q44281	Pseudomonas cepaci	
C	36	47.6	10.4	2823	1	T35233	Natural killer lyt	
C	37	47.6	10.4	12700	1	V62131	HSV-2 strain SB5 C	
C	38	47.6	10.4	11705	1	V62160	HSV-2 strain SB5 C	
C	39	47.4	10.4	1610	1	Q57657	Allele D4.7 of the	
C	40	47.4	10.4	1608	1	T27547	Recombinant human	
C	41	47.4	10.4	4425	1	V32641	Human choridin cDN	
C	42	47.4	10.4	1682	1	V62175	HSV-2 strain SB5 C	
C	43	47.2	10.3	4356	1	Q37543	Cardiac adenyllyl c	

RESULT	1
ID	T48239 standard; DNA; 1876 BP.
AC	T48239.
DT	09-APR-1997 (first entry)
DE	Neuronal nicotinic acetylcholine receptor alpha
KW	Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter; ligand-gated receptor; ds.
OS	Homo sapiens.
FN	Key
FT	cds
FT	location/Qualifiers
FT	73..1581
FT	/*tag= a
PN	W09641876-A1.
PD	27-DEC-1996.
PF	07-JUN-1996; U09775.
PR	07-JUN-1995; US-484722.
PA	(SIBI-) SIBIA NEUROSCIENCES INC.
PI	Ellicott KY, Harpold KM;
DR	WPI; 97-065463/06.
DR	P-PSDB; W09025.
PT	Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
PS	disclosure: Page 71-73; 108pp; English.
CC	A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of
CC	the human neuronal nicotinic acetylcholine receptor (nAChR). Host
CC	cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7
CC	nucleic acids, opt. in combination with other alpha and/or beta
CC	subunit nucleic acids (see also T48232-38, T48240-41), express
CC	recombinant nAChR subunits useful for identifying cpds. that
CC	modulate the activity of human nAChRs.
SO	Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;
Query Match	27.8%; Score 127; DB 1; Length 1876;
Best Local Similarity	100.0%; Pred. No. 2,2e+13;
Matches 127; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	321 GCGCCGAGCGCAGAGCCGCGGACAGCGCAGCTGAGAGCGCCGCTGCTGCACT 380
DB	1 GCGCGAGGCGCAGAGCGCCGCGGACAGCGCAGAGCTGAGAGCGCCGCTGCTGCACT 60
QY	381 CCGGAGCTCAATAGTGGCTGCGCGGAGAGGCTGCTGCGCGCTGCGCTGCTC 440
DB	61 CCGGAGCTCAATAGTGGCTGCTGCGCGGAGAGGCTGCTGCGCGCTGCGCTGCTC 120
QY	441 CTGACAG 447
DB	121 CTCACAG 127
RESULT	2
ID	V12197 standard; cDNA; 1876 BP.
AC	V12197;
DT	14-MAY-1998 (first entry)
DE	Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.
KW	Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit; brain tissue; screening; NACR; antibody; ds.
OS	Homo sapiens.
FN	Key
FT	cds
FT	location/Qualifiers
FT	73..1581
FT	/*tag= a
FT	/product= "neuronal nicotinic acetylcholine receptor
FT	alpha-7 subunit"
W09420617-A2.	
15-SEP-1994.	
DN	

PF 08-MAR-1994: U02447.
 PR 08-MAR-1993; US-028031.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KT, Ellis SB, Harpold MM;
 DR WPI: 94-303024/37.
 DR P-PSDB: W4153.
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -
 PT also transformed cells useful for screening cpds. which modulate
 PT activity of the receptor
 PS Claim 8; Page 78-79; 99pp; English.
 CC The present sequence encodes a human neuronal nicotinic acetylcholine
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta
 CC NACHR subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal NACHR.
 CC Subunit specific antibodies may be used to monitor the distribution
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes.
 CC Sequence 1876 BP; 369 A; 553 C; 530 G; 423 T;
 SQ

Query Match 27.8%; Score 127; DB 1; Length 1876;
 Best Local Similarity 100.0%; Pred. No. 2.2e-13;
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 GCGCCGAGCGGAGCGGCGGAGACGCTGAGCGCGCGGCTCCTCAGCT 380
 DB 1 GCGCCGAGCGGAGCGGCGGAGACGCTGAGCGCGCGGCTCCTCAGCT 60
 QY 381 CCGGAGCTCAATCGCTGCTCGCGGAGCGCTGCTGCGCTGCGCGCTC 440
 DB 61 CCGGAGCTCAATCGCTGCTCGCGGAGCGCTGCTGCGCTGCGCGCTC 120
 QY 441 CTGCACG 447
 DB 121 CTGCACG 127

RESULT 3
 051543/c
 051543 standard; DNA: 3861 BP.
 051543:
 24-MAY-1994 (first entry)
 DE Lipopolysaccharide induced protein gene.
 KW Macrophage; induced; lipo-polysaccharide; antitumor;
 KW antiinflammatory; trypanocidal agent; antibody; cell proliferation;
 KW activation; cytotoxicity; ds.
 OS Homo sapiens.
 PR key
 FT cds Location/Qualifiers
 FT 1810..3315
 FT /*tag- a
 FT 1980..2228
 FT /*tag- b
 FT /note- "Intron 1 is only partially sequenced
 FT (represented by (N)40 in the sequence
 FT estimated length ca. 5400 bp"
 FT 2615..2846
 FT /*tag- c
 FT /note- "Intron 2 is only partially sequenced
 FT (represented by (N)40 in the sequence
 FT estimated length ca. 7900 bp"
 FT 2907..2993
 FT /*tag- d
 FT /note- "Intron 3 is only partially sequenced
 FT (represented by (N)40 in the sequence
 FT estimated length ca. 1000 bp"
 FT Intron
 FT Intron

FT poly-a-signal 3841..3846
 FT /*tag- e
 FT W09322437-A.
 FT 11-NOV-1993.
 PD 28-APR-1993; E01022.
 PR 30-APR-1992; E010231.
 PA (INNO-) INNOGENETICS NV SA.
 PI Devos K, Franssen L, Van De Voorde A, Van Heuverswyn H;
 DR WPI: 93-368796/46.
 DR P-PSDB: R43682
 PT New polypeptide induced in macrophage(s) by lipo-polysaccharide -
 PT useful e.g. as antitumor, antiinflammatory or trypanocidal
 PT agent, also related nucleic acid, antibodies, anti-sense cpds.
 PT etc.
 PS Claim 9; Fig 1; 108pp; English.
 CC The polypeptide induced in macrophages by lipopolysaccharide
 CC stimulates cell proliferation (esp. when costimulated with IL-4)
 CC promote activation, cytotoxicity, and mobilisation of LAK cells;
 CC promote recruitment of suppressive peritoneal exudate cells;
 CC promote generation of immunocompetent lymph node cells (LNC) and
 CC have trypanocidal and trypanolytic activity. The human and murine
 CC sequences are given in (051543-45), peptide fragments able to
 CC generate antibodies are given in (R51951-61)
 CC Sequence 3861 BP; 542 A; 1241 C; 1292 G; 658 T;
 SQ

Query Match 15.1%; Score 69.2; DB 1; Length 3861;
 Best Local Similarity 53.3%; Pred. No. 0.00033;
 Matches 146; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 35 CAGCCCTTGCGCTGCGCAGAGCGCGGAGCGCGCGCTCGTGAGACTGGGGTG 94
 DB 2086 CTGCCCG 2027
 QY 95 GAGTGGCCGAGCGCTACCCAGCGCGGAGATCTCCGCTACACCTGGGCTGAGT 154
 DB 2026 GAGCGGAGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1967
 QY 155 TCCCTGGTGGCGCGCGAGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 214
 DB 1966 TGCACCGGTGCTGGAATCTACTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1907
 QY 215 GCGGGGCGGGGCGCTGACAGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 274
 DB 1906 GCACGACGCGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1847
 QY 275 CCGCGCTCTCTTAAGGCGCGCGAGCGCGCG 308
 DB 1846 CCG 1813

RESULT 4
 X53491
 ID X53491 standard; DNA: 114955 BP.
 AC X53491;
 DT 05-JUL-1999 (first entry)
 DE Human adenovase A1 receptor antisense oligonucleotide fragment.
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; palm; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma.
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metast
 OS Synthetic.
 PN W09913886-A1.
 PD 25-MAR-1999.
 PF 17-SEP-1998; U19419.
 PR 09-JUN-1998; US-093972.
 PR 17-SEP-1997; US-059160.

KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;
 KM multicystic renal dysplasia; renal agenesis; hydronephrosis;
 KM Von Mayer-Rokitansky-Kuester disorder; bifid ureter; ss.
 OS Homo sapiens.
 PH Key
 FT 5'UTR Location/Qualifiers
 FT 1..1088
 FT /*tag- a
 FT 1089..1189
 FT /*tag- b
 FT /product- "USF2 gene product"
 FT /note- "contains introns"
 FT 1089..1150
 FT /*tag- c
 FT /number- 1
 FT 1151..1435
 FT /*tag- d
 FT /number- 1
 FT 1436..1482
 FT /*tag- e
 FT /number- 2
 FT 1483..1571
 FT /*tag- f
 FT /number- 2
 FT 1572..1690
 FT /*tag- g
 FT /number- 3
 FT 1691..1792
 FT /*tag- h
 FT /number- 3
 FT 1793..1993
 FT /*tag- i
 FT /number- 4
 FT 1994..2437
 FT /*tag- j
 FT /number- 4
 FT 2438..2588
 FT /*tag- k
 FT /number- 5
 FT 2589..2708
 FT /*tag- l
 FT /number- 5
 FT 2709..2796
 FT /*tag- m
 FT /number- 6
 FT 2797..3074
 FT /*tag- n
 FT /number- 6
 FT 3075..3132
 FT /*tag- o
 FT /number- 7
 FT 3133..10631
 FT /*tag- p
 FT /number- 7
 FT 10632..10726
 FT /*tag- q
 FT /number- 8
 FT 10727..10879
 FT /*tag- r
 FT /number- 8
 FT 10880..11008
 FT /*tag- s
 FT /number- 9
 FT 11009..11099
 FT /*tag- t
 FT /number- 9
 FT 11100..11186
 FT /*tag- u
 FT /number- 10
 FT 11190..11820
 FT /*tag- v
 FT 11723..11728
 FT /*tag- w
 FT PN W09815650-A2..

PD 16-APR-1998.
 PF 09-OCT-1997; E05583.
 PR 09-OCT-1996; EP-202820.
 PA (VIAA-) VIAMS INTERNUNIVERSITAIR INST BIOTECHNOC.
 PI Flynns JPCJ Groenen PMA, Van De Ven WJM;
 DR WPI; 98-240833/21.
 DR P-PSDB; W60569.
 PT Hydronephrosis gene - useful to treat or diagnose renal diseases and
 PT disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction
 PT obstruction, multicystic renal dysplasia or renal agenesis
 PS Disclosure; Fig 6A-B; 73pp; English.
 CC This DNA encodes a USF2 gene derived from human chromosome 19. A
 CC translocation partner to this gene on chromosome 6 is the hydronephrosis
 CC gene (HNG) product. The HNG gene can be used as a starting point to
 CC design suitable compounds or techniques for the treatment of renal
 CC diseases or disorders, or nucleotide probes for diagnosing cells involved
 CC in renal diseases or disorders. A protein or a fragment encoded by HNG
 CC gene can be used as a starting point for preparing suitable antibodies
 CC for diagnosing cells involved in renal diseases and disorders. The
 CC products and method can be used to treat or diagnose renal diseases and
 CC disorders selected from vesical-ureteral reflux, uni or bilateral
 CC pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal
 CC agenesis, renal aplasia, hydronephrosis, Von Mayer-Rokitansky-Kuester
 CC disorder and bifid ureter.
 SQ Sequence 11820 BP; 745 A; 1498 C; 1443 G; 683 T;
 Query Match 12.6%; Score 57.8; DB 1; Length 11820;
 Best Local Similarity 54.6%; Pred. No. 0.018;
 Matches 183; Conservative 0; Mismatches 142; Indels 10; Gaps 3;
 QY 14 GAGGTAGAGCCTGGCCTGGGCGAGCCCTGGC--CTGGCCAAAGCGCGGACCGGAC 71
 DB 1246 GCGGCCGCTCGGAGTCATGGGGCGCCCGCGCTGACAGCCGCGGGCGCGCGGGGCGC 1187
 QY 72 CCGCTCGGTGAGACTGGGGGTGAGGTGCCGAGACGTACCAGCGCCGGAGTACTTC 131
 DB 1186 CGGGGCGCGGGGCGACAGCGCGGCGCGG-----GATCTTACTGCGGCGGCGACGCGG 1132
 QY 133 CCGCTCACACTCGGCGGTGCATCTCTGGGTGGCGCCGAGACCTGGCCGCGCTGGA 191
 DB 1131 TGGCCGAGCAGCGGGATCCAGACC---GGGTCCAGCATGTCCATGGGGGGGGGGGG 1075
 QY 132 GGGATGCGGGCGGCGGAGCGGGGCGGGGCGGGCTGTCACTGAGAGCGCGCGCGG 251
 DB 1074 CGGGGGGGCGCGGGCGCGGGGCGGGGAGGAGGGGAGGGAGGGAGGGGGGG 1015
 QY 252 GCGGGCGGGGCGGGGCGGGCGCGGGCTCTTAAAGCGCGGCGGACGGCGAGG 311
 DB 1014 GGGGCGGGGCGGGGCGGGCGGGCGGTCTACAGCCCGCGGCGGAGGAGGAGGAGG 955
 QY 312 TGCCCTGTGGCGCGAGCGAGCGCGCGGCGGACA 346
 DB 954 AGATACGGGAGCGCGCTCGCTGATCAGGGGACA 920
 RESULT 9
 Q76213/C
 ID 076213 standard; DNA; 12001 BP.
 AC 076213;
 DT 04-AUG-1995 (first entry)
 DE HSV L/ST region.
 KW HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.
 OS Herpes simplex virus type 1.
 FH Key
 FT Location/Qualifiers
 FT tata_signal
 FT 199..204
 FT /*tag- a
 FT misc_binding
 FT 220..225
 FT /*tag- b
 FT /note- "ICP4 binding site"
 FT 229
 FT misc_rna
 FT /*tag- c
 FT /note- "5' end of the L/STs"


```

FT      misc_rna      370..372
FT      /tag= d
FT      /note= "first codon of ORF-1"
PN      WO9428156-A.
PD      08-DEC-1994.
PF      20-MAY-1994; U05770.
PR      (DAND ) DANA FARBER CANCER INST INC.
PA      Schaffer PA, Yeh L.
PI      WPI: 95-022825/03.
DR      Herpes Simplex Virus (HSV) specific junction spanning transcript
PT      - for inhibiting HSV L/ST synthesis, in the treatment of HSV
PT      infection.
PS      Disclosure; Page 38-44; 64pp; English.
CC      An HSV-specific junction spanning transcript (L/ST) maps at the 5'
CC      end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,
CC      and extends into the c repeat sequences of HSV DNA. Compounds
CC      that inhibit HSV L/ST synthesis may be used as anti-HSV
CC      virucides. The HSV-1 DNA sequence in the region of the L/STs
CC      is given in Q76213.
Sequence 12001 BP: 1568 A; 4557 C; 4315 G; 1561 T;

Query Match      12.3%; Score 56.4; DB 1; Length 12001;
Best Local Similarity 50.3%; Pred. No. 0.03; Mismatches 161; Indels 1; Gaps 1;
Matches 164; Conservative 0; Mismatches 161; Indels 1; Gaps 1;

QY      43 GGCCTGGCCAGAGGCGGCGAGAGCCCGCTGCTGAGACTGGGGTGTGAGTGC 102
DB      2319 GGGAGGGGGCGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGG 2260
QY      103 CGGACGCTACCCAGCGCGGAGTACTCCGCTCACAACCTCGGGCTGCACTTCC 162
DB      2259 CGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGG 2200
QY      163 TGGCCCGCGAGACGCTGGCGCGGCTGAGAGGATGGCGGGGAGAGGGGCGG 222
DB      2199 GAGAGGGGCGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGG 2140
QY      223 GGGGCTGCTGACGTGAGAGGCGCGGGGCGGCGGGGCGGCGCGCGCTC 282
DB      2139 GAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGG 2080
QY      283 CTTAAGGCGCGGAGCGGAGCGGAGGTGCTCTGTGCCGAGCGGAGCGCGG 342
DB      2079 GAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGG 2021
QY      343 GACAGCCGAGACGTGAGCGGCGCGG 368
DB      2020 GAGGGGCGGAGGGGCGGAGGGGCGG 1995

RESULT 10
ID      V60339
AC      V60339;
DE      04-DEC-1998 (first entry)
DE      cDNA sequence of fibroblast growth factor-2 (FGF-2).
KW      Fibroblast growth factor-2; FGF-2; leaderless protein; inhibition;
KW      export; angiogenesis; restenosis; treatment; tumour; inflammation;
KW      cell proliferation; diabetes; retinopathy; infection;
KW      polycystic kidney disease; atherosclerosis; ss.
OS      Homo sapiens.
PN      WO9837880-A1.
PD      03-SEP-1998.
PF      25-FEB-1998; U03689.
PR      26-FEB-1997; US-607014.
PA      (CIBL-) CIBLEX CORP.
PI      Baird A, Florkiewicz RZ;
DR      WPI: 98-495377/42.
PT      Inhibiting export of leaderless protein with agent that inhibits
PT      binding to transporter protein - especially for treating
PT      angiogenesis and restenosis by preventing export of fibroblast

```

```

PT      growth factor, also methods for identifying leaderless proteins and
PT      their transporters
PS      Claim 2; Pages 53-54; 116pp; English.
CC      The present sequence encodes fibroblast growth factor-2 (FGF-2), a
CC      leaderless protein. A leaderless protein refers to a protein that is
CC      found in an extracellular environment, but lacks a canonical leader
CC      sequence. The specification describes a method for inhibiting export of
CC      a leaderless protein from a cell. The method comprises treating
CC      the cell with an agent that inhibits binding between the leaderless
CC      protein and a transport molecule. Treatment with the inhibiting agent
CC      is specifically used to treat angiogenesis and restenosis, i.e. where
CC      expression of FGF-2 is inhibited, and the agent is applied to endothelial
CC      or smooth muscle cells. Other applications are treatment of tumours
CC      (melanoma, teratocarcinoma, ovarian carcinoma, bladder cancer and
CC      neuroblastoma), inflammation, cell proliferation, complications of
CC      diabetes (e.g. retinopathy), viral, bacterial or fungal infections,
CC      polycystic kidney disease and atherosclerosis.
SQ      Sequence 1120 BP: 260 A; 293 C; 342 G; 225 T;

Query Match      12.3%; Score 56; DB 1; Length 1120;
Best Local Similarity 47.7%; Pred. No. 0.055;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY      41 CTGGCTGGCCAGAGGCGGCGAGAGCCCGCTGCTGGTGGAGACTGGGGTGTGAGT 100
DB      83 CGGCTTCCCAAGGCGGCGCTCCGCGGAGACACCATCGTGAACCCAGTCCCGGGCG 142
QY      101 CCCGAGACGTACCCAGGCGGCGGAGTACTCCGCTCACAACCTGGGCTGCACTTCC 160
DB      143 CGGCTGCGCGCGGAGCCAGGCGGCGGCGGAGAGAGAGGCGGAGCGGCTGAGGCTG 202
QY      161 GTGGCGCGGAGAGCGTGGCGCGGCTGAGAGGATGGCGGGGAGCGGGGCGGGG 220
DB      203 GGGAGCGGCGGCGGCGGCGGCGGCTCGGCGGCGGAGGAGGCTGGGGGCGG 262
QY      221 GCGGGGCTGCTGACGTGAGAGGCGCGCGGCGGCGGCGGCGGCGGCGCGCG 280
DB      263 CCGTGGCGGAGGCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCTCCCG 322
QY      281 TCTTAAGGCGCGGAGCGGAGCGGAGGTGCTCTGTGGCGGAGCGGAGCGCGCG 340
DB      323 CGGCTCAGCGGCTCGGGGATCCGCGCGGCGGCGGCGGCGGAGGACATGGCAGCG 382
QY      341 GCGACAGCCGAGAGCGGAGCGCGCGGCTGCTGACGCTCCG 384
DB      383 TCACACAGCTCCCGCTTGCGCCGAGAGATGGCGGAGCGGCGG 426

RESULT 11
ID      V10362/c
AC      V10362;
DE      30-JUN-1998 (first entry)
DE      Infected cell protein number 4 alpha-4 gene.
KW      Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;
KW      therapeutics; ss.
OS      Herpes simplex virus.
FH      Key      Location/Qualifiers
FT      CDS      361..4257
FT      /tag= a
FT      /product= "Infected cell protein"
PN      WO9804709-A2.
PD      05-FEB-1998.
PF      23-JUL-1997; U12904.
PR      26-JUL-1996; US-690473.
PA      (ARCH-) ARCH DEV CORP.
PI      Leopardi R, Roizman B;
DR      P-PSDB; W40200.
PT      Use of herpes simplex virus ICP4 polypeptide - useful for, e.g.
PT      blocking apoptosis in cells, production of proteins and gene therapy
PS      Disclosure; Fig 2; 63pp; English.

```


CC The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein
 CC number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block
 CC cell apoptosis. Similarly the administration of an agent that inhibits
 CC ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.
 CC This can be used for the immortalisation of cells, production of
 CC proteins, gene therapy, or inhibition of cell death induced in vivo.
 CC They can also be used for production of therapeutics comprising
 CC inhibitors of HSV ICP4 function, useful for treating HSV infection.
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 11.6%; Score 53; DB 1; Length 4257;
 Best Local Similarity 48.7%; Pred. No. 0.13;
 Matches 172; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

QY 88 GGGGGTGGAGTCCCGGAGCCGACGAGAGGAGTACCTCCGCTACACTCGGG 147
 DB 2751 GGGGGCCGGGGCGGCGCTGCGCCCGCGGCGCGCGGGCGGGCGGGGGCGGG 2692
 148 CTGCACTTCCCTGGTGGCGCGCGAGAGCTGGCCCGGGGTGAGAGAGTGGCGGG 207
 DB 2691 CTGGGGCCCGGGGGGGGTGAGAGGGGGGGCGGGCGGGAGGGGCGCGGCTCGGA 2632
 QY 208 GACGGGGGGGGGGGGCGGCTGCTGACGTGAGAGGCGCGGGGGGGGGGGGGGG 267
 DB 2631 GCGGGGGGGTCCCGCCCTCTTCTGCTTCTGGGGGTGCGGGGGCGCGCTCGGG 2572
 QY 268 GCGCGCGCGCGCTCTTAAAGCGCGAGCGGCGAGGCTCTGTTGGCGCGCA 327
 DB 2571 CGGCGGGGGCGGGCGGCGGACTCTTGGCGCTTGGCGCCCGCGGGGGGGGGCGG 2512
 QY 328 GCGGAGGCGCGGGGGGACAGCGAGCTGAGAGCGCGCGCTGCTGAGCTCGGGAGC 387
 DB 2511 GCGGGCGCGGCGGCTGCGCGGCTC-CGTTGCTGCTGCGCGCGCGCGGACAGGGGGC 2453
 QY 388 TCACATGCGCTGCTGCGCGGAGGCGTGTGCTGGCGCTGGCGCGCTGCTC 440
 DB 2452 GCAGGCTCTGTTCTCAACAGCAGAGTCCGGCGGGCGCGCGGAGCTC 2400

RESULT 12

V68520/c
 ID V68520 standard; DNA; 4257 BP.
 AC V68520;

DT 29-JAN-1999 (first entry)
 DE The nucleotide sequence of the Herpes simplex virus ICP4.
 Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;
 stimulation; inhibition; HSV infection; ss.
 US Herpes simplex virus.

FS Key Location/Qualifiers
 FH 361..4257
 FT /tag- a
 FT /product= "HSV ICP4 protein"

PN MO9846637-A2.
 PD 22-OCT-1998.
 PE 16-APR-1998; U07573.
 PR 16-APR-1997; US-843659.
 PA (ARCH-) ARCH DEV CORP.
 PI Leopardi R, Roizman B;
 DR WPI: 98-594559/50.
 DR P-PSDB: M80810.
 PT Use of herpes simplex virus U(s)3 polypeptide - for developing
 PT products for modulating apoptosis in cells and for identifying
 PT compounds which act as stimulators or inhibitors of apoptosis
 PS Example 2; Pages 60-63; 85pp; English.
 CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in
 CC and products can be used to identify compounds of apoptosis. The methods
 CC (stimulate or inhibit) apoptosis in cells. They can be used to
 CC immortalise cells for the study of these cells or for growing cells
 CC in large numbers for the productions of proteins. They can also be
 CC used for stimulating apoptosis in cells, e.g. for treating a subject
 CC with a HSV infection.

SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 11.6%; Score 53; DB 1; Length 4257;
 Best Local Similarity 48.7%; Pred. No. 0.13;
 Matches 172; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

QY 88 GGGGGTGGAGTCCCGGAGCCGACGAGAGGAGTACCTCCGCTACACTCGGG 147
 DB 2751 GGGGGCCGGGGCGGCGCTGCGCCCGCGGCGCGCGGGCGGGCGGGGGCGGG 2692
 QY 148 CTGCACTTCCCTGGTGGCGCGCGAGAGCTGGCCCGGGGTGAGAGAGTGGCGGG 207
 DB 2691 CTGGGGCCCGGGGGGGGTGAGAGGGGGGGCGGGCGGGAGGGGGCGCGGCTCGGA 2632
 QY 208 GACGGGGGGGGGGGGCGGCTGCTGACGTGAGAGGCGCGGGGGGGGGGGGGGG 267
 DB 2631 GCGGGGGGGTCCCGCCCTCTTCTGCTTCTGGGGGTGCGGGGGCGCGCTCGGG 2572
 QY 268 GCGCGCGCGCGCTCTTAAAGCGCGAGCGGCGAGGCTCTGTTGGCGCGCA 327
 DB 2571 CGGCGGGGGCGGGCGGCGGACTCTTGGCGCTTGGCGCCCGCGGGGGGGGGCGG 2512
 QY 328 GCGGAGGCGCGGGGGGACAGCGAGCTGAGAGCGCGCGCTGCTGAGCTCGGGAGC 387
 DB 2511 GCGGGCGCGGCGGCTGCGCGGCTC-CGTTGCTGCTGCGCGCGCGGACAGGGGGC 2453
 QY 388 TCACATGCGCTGCTGCGCGGAGGCGTGTGCTGGCGCTGGCGCGCTGCTC 440
 DB 2452 GCAGGCTCTGTTCTCAACAGCAGAGTCCGGCGGGCGCGCGGAGCTC 2400

RESULT 13

Q76213
 ID Q76213 standard; DNA; 12001 BP.
 AC Q76213;
 DT 04-AUG-1995 (first entry)
 DE HSV L/ST region.
 KW HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.
 OS Herpes simplex virus type 1.
 FH Key Location/Qualifiers
 FT 199..204
 FT /tag- a
 FT /note- "ICP4 binding site"

FT misc_binding 220..225
 FT /tag- b
 FT /note- "ICP4 binding site"

FT misc_rna 229
 FT /tag- c
 FT /note- "5' end of the L/STs"

FT misc_rna 370..372
 FT /tag- d
 FT /note- "first codon of ORF-1"

PN MO9428156-A.
 PD 08-DEC-1994.
 PE 20-MAY-1994; U05770.
 PR 20-MAY-1993; US-065146.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Schaffer PA, Yeh L;
 DR WPI: 95-022825/03.
 DR Herpes Simplex Virus (HSV) specific junction spanning transcript
 PT - for inhibiting HSV L/ST synthesis, in the treatment of HSV
 PT infection.
 PS Disclosure; Page 38-44; 64pp; English.
 CC An HSV-specific junction-spanning transcript (L/ST) maps at the 5'
 CC end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,
 CC and extends into the c repeat sequences of HSV DNA. Compounds
 CC that inhibit HSV L/ST synthesis may be used as anti-HSV
 CC virucides. The HSV-1 DNA sequence in the region of the L/STs
 CC is given in Q76213.
 SQ Sequence 12001 BP; 1568 A; 4557 C; 4315 G; 1561 T;

Query Match

11.5%; Score 52.4; DB 1; Length 12001;

RESULT 15

ID 047927/c
O47927 standard; cDNA; 4403 BP.

AC 24-MAR-1994 (first entry)
DE Paired basic amino acid converting enzyme (PACE) 4 gene;
KM hypertension; blood; coagulation; ss.
OS Homo sapiens.
FH Key location/Qualifiers
FT 170..3079
FT cds /tag= a
FT /product= PACE 4.

PN MO9318159-A.
PD 16-SEP-1993.
PE 09-MAR-1993; U02147.
PR 09-MAR-1992; US-848629.
PA (CHIR) CHIRON CORP.
PI Barr PJ, Klefer MC;
DR WPI: 93-303473/38.
DR P-PSNB; RA1662.
PT New polynucleotide encoding PACE endo-peptidase - used for e.g.
reducing blood pressure
PT Claim 1; Figure 1; 81pp; English.
CC The mammalian endopeptidases PACE 4 and PACE 4.1 (an alternative
form of PACE 4) are involved in the production of mature
polypeptides from precursor polypeptides by cleavage at pairs of
basic amino acids e.g Lys-Arg, Lys-Lys and Arg-Arg. The enzymes
are stimulated by the presence of calcium ions and inhibited by
phenylmethyl sulphonyl fluoride. PACE 4 and 4.1 convert proteinin to
renin in the human kidney, causing elevation of blood pressure.
CC They can therefore be used to identify other inhibitors of their
action and may also be used to inhibit blood coagulation.
SQ Sequence 4403 BP; 1016 A; 1253 C; 1214 G; 920 T;

Query Match 11.3%; Score 51.8; DB 1; Length 4403;
Best Local Similarity 49.2%; Pred. NO. 0.2; Indels 2; Gaps
Matches 164; Conservative 0; Mismatches 167;

Dy 141 GCATCTCCGGGCGGCCGCCCATTCGGGCAAGCGCGGGGTCCACGAGCCCCGGCGGC 82
Qy 363 CGCAGCGCTGCTGCTCACCTCCGGACTCAACATGCTCCTGCGCGGAGGCTC 416
Db 81 CAGGCGGGGTCCGAGGGTGGCGCTCCCTCGGCGCTGTGACACGGCACTTTC 28

Dy 53 GAGGCGCGAGCGCCGAGACCCTCGTGAGACTGGGGGTGAGTGTCCCGAGGCTAC 112
Qy 373 GGCGCGCGCGCGGGGCGCGGAGCAGCGCGGAGCGCCAGCAGACCGCCGA 314
Db 113 CCAGCGCGCGGAGTACTCTCCGCTCACAACCTCGGCTCAATTCCCTGGTGGCGCGCA 172
Qy 313 GGGACCGCGCGGAGCGGCGCGGAAACCGCGGCCCTCGCGCGCCCGCCCGCGCC 254
Db 173 GAGCTGCGCGGCTTGAGAGGATGGCGGGGCGGGAGCGGGGCGCGGCTGCT 232
Qy 253 CCCCGCGCGCGCGGTGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 194
Db 233 ACGTGAGAGAGCGCGGGGCGGGGCGGGGCGGGGCGCGCGCGCGCGCGCTTTAAGGCG 292
Qy 193 CGCAGGCGCGCGCGCGGAGGAGCATACGGGAGACAGGCTGCG--CGCGCCCGAGCTGCG 136
Db 293 CGCAGCGAGCGCGCGCGAGGTGCTCTGTGAGCGCCCAAGCGCGAGCGCGCGAG 352
Qy 135 AGTGGCGCGGGGGGTGAGTGTCCCTTTTAAGCCGCTCCGCGCGGGGAAGCGCGCG 76
Db 353 ACGTGAAGCGCGCGCGCTCGGTGAGCTCCGG 385
Dy 75 GCCGCGCGCACCGCGCGCGCCACACACGCGCG 43

Sun Sep 19 10:27:05 1999

us-08-956-518a-94.rng

Page 9

Search completed: September 18, 1999, 00:34:06
Job time: 19006 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:07; Search time 2825.05 seconds
(without alignments)
319.091 Million cell updates/sec

Title: US-08-956-518a-94
Perfect score: 457
Sequence: 1 AGAAGCAGAGGAGAGGAGTAG.....CTCCGACAGGTAAGCCAC 457

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database: ESR:**

1: em_est1:.*
2: em_est2:.*
3: em_est3:.*
4: em_est4:.*
5: em_est5:.*
6: em_est6:.*
7: em_est7:.*
8: em_est8:.*
9: em_est9:.*
10: em_est10:.*
11: em_est11:.*
12: em_est12:.*
13: em_est13:.*
14: em_est14:.*
15: em_est15:.*
16: em_est16:.*
17: em_est17:.*
18: em_est18:.*
19: em_est19:.*
20: gb_est1:.*
21: gb_est2:.*
22: gb_est3:.*
23: gb_est4:.*
24: gb_est5:.*
25: gb_est6:.*
26: gb_est7:.*
27: gb_est8:.*
28: gb_est9:.*
29: gb_est10:.*
30: gb_est11:.*
31: gb_est12:.*
32: gb_est13:.*
33: gb_est14:.*
34: gb_est15:.*
35: gb_est16:.*
36: gb_est17:.*
37: gb_est18:.*
38: gb_est19:.*
39: gb_est20:.*
40: gb_est21:.*
41: gb_est22:.*
42: gb_est23:.*
43: gb_est24:.*
44: gb_est25:.*
45: gb_est26:.*
46: gb_est27:.*
47: gb_est28:.*
48: gb_est29:.*
49: gb_est30:.*
50: gb_est31:.*
51: gb_est32:.*
52: em_est20:.*
53: em_est21:.*

54: em_est22:.*
55: em_est23:.*
56: em_est24:.*
57: em_est25:.*
58: em_est26:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	77.4	16.9	641	45	A1357868	A1357868 qv13b02.x
C 2	74.2	16.2	430	45	A1367590	A1367590 qg30a12.x
C 3	73	16.0	759	26	W28277	W28277 4498 Human
C 4	68.2	14.9	459	44	A1285778	A1285778 qv5b02.x
C 5	67.8	14.8	1461	25	W15723	W15723 mb53c01.r1
C 6	65.8	14.4	457	47	A1469170	A1469170 ta19d04.x
C 7	65	14.2	547	47	A1525191	A1525191 promrna-9
C 8	63	13.8	635	47	A1525004	A1525004 promrna-4
C 9	61.6	13.5	1208	26	W29297	W29297 mb99e03.r1
C 10	61.4	13.4	1075	26	W29419	W29419 mb98h11.r1
C 11	61	13.3	1171	26	W29436	W29436 mb99g07.r1
C 12	60.4	13.2	453	27	AA008005	AA008005 mg64c12.x
C 13	60.2	13.2	1161	26	W41959	W41959 mc68b10.r1
C 14	60.2	13.2	403	47	A1468375	A1468375 tg58d01.x
C 15	59.8	13.1	859	26	W28841	W28841 52d7 Human
C 16	59.6	13.0	1681	49	A1653819	A1653819 ty01h01.x
C 17	59.4	13.0	1243	26	W42204	W42204 mc96d11.r1
C 18	59.2	13.0	1242	26	W29459	W29459 mc03f02.r1
C 19	59.2	13.0	1145	27	AA028476	AA028476 m12le10.r
C 20	59.2	13.0	362	42	A1150327	A1150327 qf35a01.x
C 21	59.2	13.0	407	44	A1250825	A1250825 gx04f05.x
C 22	59	12.9	1311	25	W15735	W15735 mb53g03.r1
C 23	59	12.9	1084	27	AA028875	AA028875 mh90d12.x
C 24	59	12.9	1257	27	W98479	W98479 mg20c12.r1
C 25	58.8	12.9	1138	45	A1374006	A1374006 SMOVAFCAP
C 26	58.6	12.8	1162	26	W42215	W42215 mc69h10.r1
C 27	58.6	12.8	454	50	A1700029	A1700029 tt36d12.x
C 28	58	12.7	1291	25	W15737	W15737 mb53h03.r1
C 29	58	12.7	1111	26	W41963	W41963 mc68d10.r1
C 30	58	12.7	1056	26	W42205	W42205 mc69e09.r1
C 31	58	12.7	1087	27	AA023473	AA023473 mh74e06.x
C 32	58	12.7	454	45	A1340211	A1340211 q079f12.x
C 33	58	12.7	292	47	A1500717	A1500717 tn94b08.x
C 34	57.8	12.6	1136	25	W15902	W15902 mb56e10.r1
C 35	57.8	12.6	832	26	W28836	W28836 52d2 Human
C 36	57.8	12.6	418	48	A1611745	A1611745 tu77h12.x
C 37	57.4	12.6	1049	26	W42199	W42199 mc69c09.r1
C 38	57.2	12.5	743	26	W25793	W25793 13e7 Human
C 39	57	12.5	1222	26	W42209	W42209 mc69f11.r1
C 40	56.8	12.4	415	45	A1363984	A1363984 qv35b02.x
C 41	56.8	12.4	1104	26	W41968	W41968 mc68f11.r1
C 42	56.6	12.4	1286	26	W42216	W42216 mc69h11.r1
C 43	56.6	12.4	435	44	A1302382	A1302382 gpa49h12.x
C 44	56.2	12.3	1088	25	W15897	W15897 mb56b09.r1
C 45	56.2	12.3	829	26	W28819	W28819 52b8 Human

ALIGNMENTS

RESULT 1
LOCUS A1357868 641 bp mRNA
DEFINITION qv13b02.x1 NCI-GAP K168 Homo sapiens cDNA clone IMAGE:1981419 3'
ACCESSION A1357868
NTD g4109489
VERSION A1357868.1 GI:4109489

15-FEB-1999
Similar to TR:069340 069340 ORF1, ORF2, AND ORF3. ;, mRNA sequence.

KEYWORDS	EST.
SOURCE	human
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 641)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) On Aug 21, 1998 this sequence version replaced.
BASE COUNT	34 a 421 c 131 g 24 t 31 others
ORIGIN	
FEATURES	
SOURCE	Insert Length: 1681 Std Error: 0.00 Seq primer: -40UP from Glbpco High quality sequence stop: 90. Location/Qualifiers 1..641 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1981419" /clone_lib="NCI-CGAP-Kid8" /tissue_type="renal cell tumor" /lab_host="DH10B" /note="Organ: kidney; Vector: PCMW-SPO6; Site 1: Salt; Site 2: NotI; Cloned unidirectionally. Primer: 0190 dt. Average insert size 1.2 kb. Life Technologies catalog #: 11524-014"

[illegible]

FEATURES	source
Db	225 GGGGGGGGGGGGNG 210
RESULT	2
LOCUS	AI367590/c
DEFINITION	AI367590 430 bp mRNA
ACCESION	AI367590
NID	AI367590
VERSION	AI367590.1
KEYWORDS	GI:4137335
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Euthelia; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (Bases 1 to 430)
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
CONTACT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.lml.gov) for further information. Insert Length: 276 Std Error: 0.00 Seq primer: -400p from Glbco. Location/Qualifiers 1..430

[illegible]

283 CTAAAGCGCGGAGACCCGAGCGGAGTGTGTGCCCCAGGCGCGCGG 342
Db 188 CGGG 129
QY 343 GACAGCGGAGACGTGAGCGCGCGCGCGCTGCTG 375
Db 128 GCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 96

RESULT 3
W28277 759 bp mRNA EST 08-MAY-1996
LOCUS 4498 Human retina cDNA randomly primed sublibrary Homo sapiens
DEFINITION CDNA, mRNA sequence.
ACCESSION W28277
NID 91308243
VERSION W28277.1 GI:1308243
YWORDS EST.
ORCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 759)
AUTHORS Macke, J., Smallwood, P. and Nathans, J.
JOURNAL Adult Human Retina cDNA
COMMENT On May 5, 1995 this sequence version replaced gi:179795.

FEATURES
source
Contact: Dr. Jeremy Nathans
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics
Johns Hopkins School of Medicine
725 North Wolfe Street, Baltimore, MD 21205
Tel: 410 955 4678
Fax: 410 614 0827
Email: jeremy.nathans@jhu.edu
Clones from this library are NOT available.
PCR Primers
FORWARD: CTTTGGAGCAAGTCAGCCTGGTTAAGT
BACKWARD: GAGGTGCTTATGATATCTTCCAGGGTAA
Seq primer: GCGTAAAGCAAGAAAGATT.
Location/Qualifiers
1. 759
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone_lib="Human retina cDNA randomly primed sublibrary"
/sex="mixed (males and females)"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="E. coli strain K802"
/note="Organ: eye; Vector: lambda gt10; Site_1: EcoRI;
Site_2: EcoRI; The library used for sequencing was a
sublibrary derived from a human retina cDNA library.
Inserts from retina cDNA library DNA were isolated,
randomly primed, PCR amplified, size-selected,
into lambda gt10. Individual plaques were arrayed and
used as templates for PCR amplification, and these PCR
products were used for sequencing."

BASE COUNT 38 a 71 c 494 g 16 t 140 others
ORIGIN

Query Match 16.0%; Score 73; DB 26; Length 759;
Best Local Similarity 42.6%; Pred. No. 0.0073;
Matches 181; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 10 GGGAGAGTAGAGCCTGCTGGAGCCCGCTGGCCAGAGCGGAGGCGGAGA 69
Db 313 GGG 372
QY 70 GCGCGCTCGTGAAGACTGGGGGTGGTCCCGAGCGCTACCCAGCGCGGAGTACC 129

Db 373 GGGGCGGG 432
QY 130 TCCCGCTACACCTCGGGGTGTCAGTTCCCTGGGTGGCCCGCAGACGCTGCCGGGCTG 189
Db 433 GGGGCGGG 492
QY 190 GAGGATGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 249
Db 493 CCGGG 552
QY 250 GGGGCGGG 309
Db 553 GGGGCGGG 612
QY 310 GGTGCTCTGTGGCGCGCAGAGCGGCGGCGCGCTCTTAAAGCGCGCGAGCGCGCGCA 369
Db 613 GGGGCGGG 672
QY 370 TCGCTCAGCTCCGGGACTCAACATGCGCTGCTGCGCGGAGGCGTGTGGCTGGCGCTGG 429
Db 673 CGGNGCGGG 732
QY 430 CCGCG 434
Db 733 GGGGG 737

RESULT 4
A1285778 459 bp mRNA EST 03-FEB-1999
LOCUS 9156D02.x1 NCI-CGAP Lym6 Homo sapiens cDNA clone IMAGE:1968747 3'
DEFINITION similar to gb:X59268 TRANSCRIPTION INITIATION FACTOR IIB (HUMAN);,
mRNA sequence.
ACCESSION A1285778
NID 93924011
VERSION A1285778.1 GI:3924011
YWORDS EST.
ORCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 459)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 17, 1998 this sequence version replaced gi:2043524.

FEATURES
source
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Mark Raffeld, M.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1167 Std Error: 0.00
Seq primer: -400P from gibco
High quality sequence stop: 458.
Location/Qualifiers
1. 459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1968747"
/clone_lib="NCI-CGAP_Lym6"
/tissue_type="mantle cell lymphoma"
/lab_host="SOLR (Stratagene, kanamycin resistant)"
/note="Organ: lymph node; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
oligo dt. Average insert size 0.8 kb. Non-amplified

BASE COUNT	ORIGIN
24 a	library, -5' adaptor sequence: 5' GATTTCGGACACAG 3' -3'
335 c	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'
38 g	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'
37 t	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'
25 others	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'

Tissue Procurement: Mark Ratfeld, M.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bhrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 455.

FEATURES
 source
 1. 457
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2044519"
 /clone_1ib="NCI CGAP_Lym5"
 /tissue_type="follicular lymphoma"
 /lab_host="GOLR (Stratagene, kanamycin resistant)"
 /note="Organ: lymph node; Vector: pBluescript SK-; Site: 1:
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 0190 dt. Average insert size 1.2 kb. Non-amplified
 library. -5' adaptor sequence: 5' GAATTCGACAGAG 3' -3'
 adaptor sequence: 5' CTCGATTTTCTTTTCTTTT 3'."
 BASE COUNT 26 a 309 c 51 g 20 t 51 others
 ORIGIN

Query Match 14.4%; Score 65.8; DB 47; Length 457;
 Best Local Similarity 43.1%; Pred. No. 0.084;

Matches 169; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

OY 55 GCGCGGAGCCAGACGCCGCTGCTGAGAGAGTGGGAGAGTCCCGAGACGCTACCC 114
 DB 456 GCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 397
 OY 115 ACCGCCGGAGTACCTCCCGCTCACACCTCGGGCTCACTTCCCTGGTGCCGCCGAGA 174
 DB 396 GGG 337
 OY 175 CCGTGGCCCGGCTGAGAGATGCGGCGGCGGAGCGGGGGCGGGGCGTGTCTAC 234
 DB 336 GGG 277
 OY 235 GTGAGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCTTAAAGCGCG 294
 DB 276 GGG 217
 OY 295 CGAGCCGAGCGGCGAGTCTCTGTGGCCGACAGCGCCCGGGCGACAGCCGAGAC 354
 DB 216 GGG 157
 OY 355 GTGAGAGCGCGCGGCTGCTGCGAGCTCAACTGCGTGTCTCCCGCGAGGCGG 414
 DB 156 GGG 97
 OY 415 TCTGGCTGGCGCTGCGCGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 446
 DB 96 GGGGCGNNCGGGGGGCGACCCCATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 65

RESULT 7
 LOCUS A1525191 547 bp mRNA EST 18-MAR-1999
 DEFINITION Promina-9, G12.r bvtumor Homo sapiens CDNA 5', mRNA sequence.
 ACCESSION A1525191
 NID 94439326
 VERSION A1525191.1 GI:4439326
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 547)

AUTHORS
 Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun
 Yu, J., and Hood, L.
 TITLE
 Journal
 COMMENT
 On May 18, 1998 this sequence version replaced gi:3136577.

Contact: Guyang Matthew Huang
 Leroy Hood
 University of Washington
 Department of Molecular Biotechnology, Box 357730, University of
 Washington, Seattle, WA 98195
 Tel: 5106280100
 Fax: 5106280108
 Email: huanggm@yahoo.com.

FEATURES
 source
 1. 547
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="21"
 /clone_1ib="bvtumor"
 /note="Organ: Prostate; Vector: pBluescript; Directional
 CDNA library was constructed using Lambda ZP II kit
 (Stratagene). mRNA was extracted from a fresh prostate
 tumor tissues (Urology Department, University of
 Washington)."
 BASE COUNT 83 a 255 c 77 g 73 t 59 others
 ORIGIN

Query Match 14.2%; Score 65; DB 47; Length 547;
 Best Local Similarity 54.1%; Pred. No. 0.11;

Matches 113; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 160 GGGTGGCCCGCCAGACGCTGCGCCCGCTGAGAGATGCGGCGGCGGAGCGGCGG 219
 DB 508 GGG 449
 OY 220 GCGGGGCTGCTCACAGTGGAGAGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGG 279
 DB 448 GGG 389
 OY 280 CTCTTAAAGCGCGCGAGCCGAGCGGCGAGTGTCTGTGGCCCGACAGCGCCG 339
 DB 388 NCGGG 329
 OY 340 GCGGACGCGAGAGAGTGTGAGCGCGCGCGG 368
 DB 328 GGGGGCGGCGCGGCGGCGGAGGCGGAGCCG 300

RESULT 8
 LOCUS A1525004 635 bp mRNA EST 18-MAR-1999
 DEFINITION Promina-4, B11.r bvtumor Homo sapiens CDNA 5', mRNA sequence.
 ACCESSION A1525004
 NID 94439139
 VERSION A1525004.1 GI:4439139
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 635)
 AUTHORS Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun
 Yu, J., and Hood, L.
 TITLE
 Journal
 COMMENT
 On May 7, 1998 this sequence version replaced gi:3121673.

Contact: Guyang Matthew Huang
 Leroy Hood
 University of Washington
 Department of Molecular Biotechnology, Box 357730, University of

Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES	Location/Qualifiers
source	1. .635

BASE COUNT	75 a	69 c	312 g	94 t	85 others
ORIGIN					

Query Match	13.88;	Score 63;	DB 47;	Length 635;
1st Local Similarity	43.98;	Pred. No. 0.21;		
Matches 144;	Conservative	0;	Mismatches 184;	Indels 0;
			Gaps	0

QY	16	GGTAGAGACCTTGCCCTTGGGGACGCCCTTGCCCTGGCCAGAGAGCGGAGAGCCGAGAGGCCGC	75
Db	206	GGGAGNCANGAAATTATTTGGTATGAGNCCGNNGTGGGNGNNGANGANGANCCGGGNNCTTC	2655
QY	76	TCGGTGTAGACTGTGGGGGTGGAGGTGCCCGGAGCCTTACCAGCGCGGAGAGTACTCCGC	1353
Db	266	AGCAAGAGCNGCNGGGGGNCANGAGAGNNGMCNCGNNGNNGGGCCNGGGGTTAAAGNGGGG	3255
QY	136	TCACACCTCGGGCTGCAGTTCCCTGGGGTGGCCGCCGAGAGCCTGGCCCGGGCTGAGAGGA	1953
Db	326	GGNNGCNGGGGNCNGCNGAGAGGGGTGGGNGNGGCCNNGTCCGGGGCAGGGGGGGGGCGCG	3855
QY	196	TGGCGGGGGCGGAGCGGGGGCGGGGGCGGGGCTGTACGTGTGAGAGCGCGCGGGGGCG	2255
Db	386	GGGGGGGGGGGCGGCGGGGAGAGGGGGGGGGGGGGCGCCCTTGGGGGGGGGCGCNCNGCGGGGG	4455
QY	256	GGCGGGGGGGGGGCGGCCGCCCTCTCTTAAAGGCCGAGCGCCGAGCCGAGCGGAGGTGCC	3155
Db	446	GGTGTGNGNGNGGGGGCGGGTGTAGNGGGGGGCGNCGAGGNGAGGAGGGGGGGGGGGG	5055
QY	316	TCTGTGGCCGAGGCGGAGGCCCGCGGGG	343
Db	506	TGGGNGGGGGGGGGGGGTGTGGGGAG	533

Accession	LOCUS	DEFINITION	1208 bp	Mus musculus CDNA clone	EST	11-SEP-1996
U01977	W29297	mb9503.r1 Soares mouse pJNM619.5				
		IMAGE:337564 5' similar to gb:J03750.rna2				
		Mouse single stranded binding protein p9 mRNA, complete (MOUSE);				
		mRNA sequence.				

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	unpublished (1996)
COMMENT	On Sep 1, 1995 this sequence version replaced

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:218964
Seq primer: EMP1mer
High quality sequence stop: 388.
Location/Qualifiers
1..1208

	a	c	g	t	others
BASE COUNT	253	162	519	223	51
ORIGIN					

Query Match	13.5%;	Score 61.6;	DB 26;	Length 1208;
Best Local Similarity	47.7%;	Pred. No. 0.33;		
Matches 106; Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0

[illegible]

RESULT	10					11-SEP-1996
Locus	W29419	M29419	1075 bp	mRNA	EST	
DEFINITION		m89bh1..r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone				
		IMAGE:337509 5', mRNA sequence.				
ACCESSION		W29419				
NID		G1309566				
VERSION		W29419.1	GI:1309566			
KEYWORDS		EST.				
SOURCE		house mouse.				
ORGANISM		Mus musculus				

REFERENCE
AUTHORS
1 (Pages 1 to 1075)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyle, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:02:31; Search time 1962.17 Seconds
(without alignments)
499.210 Million cell updates/sec

Title: US-08-956-518a-95
Perfect score: 308
Sequence: 1 CAGCGCGCCACATAGCTCCC.....CAACATTAGGTGAGTCGCC 308

Scoring table: IDENTITY_NTC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank: 1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pl4.*
11: gb_pl5.*
12: gb_pl6.*
13: gb_pl7.*
14: gb_pl8.*
15: gb_pl9.*
16: gb_pl10.*
17: gb_pl11.*
18: gb_pl12.*
19: gb_pl13.*
20: gb_pl14.*
21: gb_pl15.*
22: gb_pl16.*
23: gb_pl17.*
24: gb_pl18.*
25: gb_pl19.*
26: gb_pl20.*
27: gb_pl21.*
28: gb_pl22.*
29: gb_pl23.*
30: gb_pl24.*
31: gb_pl25.*
32: gb_pl26.*
33: gb_pl27.*
34: gb_pl28.*
35: gb_pl29.*
36: gb_pl30.*
37: gb_pl31.*
38: gb_pl32.*
39: gb_pl33.*
40: gb_pl34.*
41: gb_pl35.*
42: gb_pl36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	275.8	89.5	641	11	AF029838	AF029838 Homo sapi

2	275.8	89.5	641	11	AF029838	AF029838 Homo sapi
3	205.2	66.8	291	9 <th>HS2E10F</th> <th>258126 H.sapiens C</th>	HS2E10F	258126 H.sapiens C
4	181.2	58.6	298	9 <th>HS2E10F</th> <th>260646 H.sapiens C</th>	HS2E10F	260646 H.sapiens C
5	181.2	58.8	137699	11 <th>AC005630</th> <th>AC005630 Homo sapi</th>	AC005630	AC005630 Homo sapi
6	154.8	50.3	192439	35 <th>AC007562</th> <th>AC007562 Homo sapi</th>	AC007562	AC007562 Homo sapi
7	146.4	47.5	178307	35 <th>AC006328</th> <th>AC006328 Homo sapi</th>	AC006328	AC006328 Homo sapi
8	87.4	28.4	5281	9 <th>HS008191</th> <th>U08191 Human R kap</th>	HS008191	U08191 Human R kap
9	51	16.6	1700	7 <th>OSNR112</th> <th>X15819 Rice ntai g</th>	OSNR112	X15819 Rice ntai g
10	50.4	16.4	171	9 <th>HS165HF</th> <th>257156 H.sapiens C</th>	HS165HF	257156 H.sapiens C
11	49.8	16.2	180	9 <th>HS11532F</th> <th>254415 H.sapiens C</th>	HS11532F	254415 H.sapiens C
12	49.4	16.0	189	9 <th>HS158R</th> <th>261392 H.sapiens C</th>	HS158R	261392 H.sapiens C
13	45	14.6	142123	11 <th>AC004690</th> <th>AC004690 Homo sapi</th>	AC004690	AC004690 Homo sapi
14	40.8	13.2	2837	10 <th>HS0G16P5</th> <th>AF014839 Frankla s</th>	HS0G16P5	AF014839 Frankla s
15	39.6	12.9	8546	2 <th>AF014839</th> <th>AF014839 Frankla s</th>	AF014839	AF014839 Frankla s
16	39.4	12.8	2991	8 <th>AF036891</th> <th>AF036891 Zee mays</th>	AF036891	AF036891 Zee mays
17	39.2	12.7	182756	9 <th>AC007789</th> <th>AC007789 Oryza sat</th>	AC007789	AC007789 Oryza sat
18	39	12.7	795	9 <th>HUMANX01</th> <th>L20600 Human annex</th>	HUMANX01	L20600 Human annex
19	39	12.7	207466	35 <th>AC007249</th> <th>AC007249 Homo sapi</th>	AC007249	AC007249 Homo sapi
20	38.8	12.6	2672	3 <th>RABTHBP</th> <th>J05602 Rabbit mult</th>	RABTHBP	J05602 Rabbit mult
21	38.8	12.6	131649	11 <th>AC005875</th> <th>AC005875 Homo sapi</th>	AC005875	AC005875 Homo sapi
22	38.4	12.5	3459	10 <th>HSNA012582</th> <th>AT012582 Homo sapi</th>	HSNA012582	AT012582 Homo sapi
23	38.4	12.5	785	10 <th>HSNA133727</th> <th>AJ133727 Homo sapi</th>	HSNA133727	AJ133727 Homo sapi
24	38.2	12.4	2992	5 <th>A63308</th> <th>A63308 Sequence 1</th>	A63308	A63308 Sequence 1
25	38.2	12.4	3289	10 <th>HSMFH1</th> <th>Y08223 H.sapiens M</th>	HSMFH1	Y08223 H.sapiens M
26	37.8	12.3	189356	11 <th>AC005746</th> <th>AC005746 Homo sapi</th>	AC005746	AC005746 Homo sapi
27	37.6	12.2	152141	10 <th>HSU07000</th> <th>U07000 Human break</th>	HSU07000	U07000 Human break
28	37.6	12.2	4510	12 <th>XMDRES9R</th> <th>Y08922 M.musculus</th>	XMDRES9R	Y08922 M.musculus
29	37.4	12.1	9373	9 <th>HUMODCA1</th> <th>M81740 Human ornit</th>	HUMODCA1	M81740 Human ornit
30	37.4	12.1	1111	9 <th>HUMODCA1</th> <th>M34157 Human ornit</th>	HUMODCA1	M34157 Human ornit
31	37.4	12.1	8042	9 <th>HUMODCA1</th> <th>M31061 Human ornit</th>	HUMODCA1	M31061 Human ornit
32	37.4	12.1	129327	10 <th>HS171N11</th> <th>AL031433 Human DNA</th>	HS171N11	AL031433 Human DNA
33	37.4	12.1	9043	10 <th>HSODCG</th> <th>X16277 Human gene</th>	HSODCG	X16277 Human gene
34	37.4	12.1	2858	3 <th>S71124</th> <th>S71124 ornithine d</th>	S71124	S71124 ornithine d
35	37.2	12.1	1919	3 <th>BTU81159</th> <th>U81159 Bos taurus</th>	BTU81159	U81159 Bos taurus
36	37.2	12.1	2122	12 <th>RNU05989</th> <th>U05989 Rattus norv</th>	RNU05989	U05989 Rattus norv
37	36.8	11.9	6945	2 <th>AF047554</th> <th>AF047554 Myxococcu</th>	AF047554	AF047554 Myxococcu
38	36.8	11.9	42999	10 <th>HSU13369</th> <th>U13369 Human ribos</th>	HSU13369	U13369 Human ribos
39	36.8	11.9	152261	17 <th>HE1CG</th> <th>X14113 Herpes simp</th>	HE1CG	X14113 Herpes simp
40	36.8	11.9	12979	17 <th>HEHSV1SU</th> <th>X02138 Herpes simp</th>	HEHSV1SU	X02138 Herpes simp
41	36.8	11.9	26245	17 <th>HS1US</th> <th>L00036 HSV1 (strai</th>	HS1US	L00036 HSV1 (strai
42	36.8	11.9	648	17 <th>S62895</th> <th>S62895 US8, US8.5-</th>	S62895	S62895 US8, US8.5-
43	36.8	11.9	135360	35 <th>AC006435</th> <th>AC006435 Homo sapi</th>	AC006435	AC006435 Homo sapi
44	36.8	11.9	177707	35 <th>AC006450</th> <th>AC006450 Homo sapi</th>	AC006450	AC006450 Homo sapi
45	36.6	11.9	6183	1 <th>STMORFD</th> <th>D45246 Streptomyce</th>	STMORFD	D45246 Streptomyce

ALIGNMENTS

RESULT 1
LOCUS AF029838 641 bp mRNA 16-DEC-1998
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
ACCESSION AF029838
NID 93757793
VERSION AF029838.1 GI:3757793
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,
Buiting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
TITLE Genomic organization and partial duplication of the human alpha7
neural nicotinic acetylcholine receptor gene
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 641)
AUTHORS Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
Breese,C., Davis,A., Hopkins,J. and Freedman,R.
TITLE Direct Submission

JOURNAL Submitted (15-OCT-1997) **Psychiatry** University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES Location/Qualifiers
SOURCE 1. 641

FEATURES
Source

Source

```

/misc.feature
/organism="Homo sapiens"
/tx.xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CRNA7-DR1"
/note="RACE product A/B/C/D"
<1..-641

```

BASE COUNT	151 a	192 c	163 g	135 t
------------	-------	-------	-------	-------

ORIGIN

Very Match	89.5%	Score 275.8;	DB 11;	Length 641;
Local Similarity	98.7%	Pred. No. 2.1e-48;		
Matches 299; Conservative	0;	Mismatches 2;	Indels 2;	Gaps 2

QY	1	CAACCCGACACATAGCTCCGCGCCCAAGCTCCGGTCCCGCCCTTGCCATTTTCCACCCGG-T	59
Db	1	CAGGCCCCACATAGCTCCCGCCCAAGCTCCGGTCCCGCCCTTGCCATTTTCCACCCGGCT	60
QY	60	CCACAGAGGGTACAGCGCGCGGGGAGAGGTGGAGCCGAGAGCTGGCGGGGGGCCCG	119
Db	61	CCACAGAGGGTACAGCGCGCGGGGAGAGGTGGAGCCGCGAGACTGGCGGGGGGCCCG	120
QY	120	CGTGGTGGCGCGGCCATGACAGCGGCTCCGGACTGGCTCTTTTCCGCGCCCTCCGC	179
Db	121	CGTGGTGGCGCGGCCATGACAGCGGCTCCGGACTGGCTCTTTTCCGCGCCCTCCGC	180
QY	180	CGGAGGTGAGGGGAGATGTCATGTCAGGGTTCAGGCCCAACCGAAGTTACTGGCCCT	239
Db	181	CGGAGGTGAGGGGAGATGTCATGTCAGGGTTCAGGCCCAACCGAAGTTACTGG-CCCT	239
QY	240	CTATCTTCCAGGAGAACACAGAGCCACAGCCGGGCTCACGCCCAACCGCACATTAAAG	299
Db	240	CTATCTTCCAGGAGAACACAGAGCCACAGCCGGGCTCACGCCCAACCGCACATTAAAG	299
QY	300	TGA	302
Db	300	TTA	302

	A.	JT	2	
	9839			
LOCUS				
DEFINITION	AF029839	689 bp	mRNA	PRI
ACCESSION	Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.			16-DEC-1998
NID	AF029839			
VERSION	93757794			
KEYWORDS	AF029839.1	GI:3757794		
SOURCE				
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
	Eutheria; Primates; Catarrhini; Homidae; Homo.			
REFERENCE	1 (bases 1 to 689)			
AUTHORS	Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J., Moore,T., Jacobs,S., Merlvetter,J., Choi,M.J., Kim,E.J., Walton,K., Bulting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.			
TITLE	Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene			
JOURNAL	Genomics 52 (2), 173-185 (1998)			
MEDLINE	99000837			
REFERENCE	2 (bases 1 to 689)			
AUTHORS	Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Breese,C., Davis,A., Hopkins,J. and Freedman,R.			
TITLE	Direct submission			
JOURNAL	Submitted (15-OCT-1997) Psychiatry University of Colorado Health Sciences Center 4200 E 9th Ave., Denver, CO 80262, USA			
FEATURES	Location/Qualifiers			

```

source
1. .689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
<1. .>689
/note="alpha 7 neuronal nicotinic receptor mRNA; used in
alternative splicing"
misc_feature
BASE COUNT      142 a      210 c      196 g      141 t
ORIGIN

```

BASE COUNT ORIGIN	142 a	210 c	196 g	141 f
----------------------	-------	-------	-------	-------

ORIGIN

Query Match	89.5%;	Score 275.8;	DB 11;	Length 689;
Best Local Similarity	98.7%;	Pred. No. 2.1e-48;		
Matches 299; Conservative	0;	Mismatches 2;	Indels 2;	Gaps 2;

QY	1	CAGGCGCCACATGATGCTCCCGCCAAAGCTCCCGGGCCCGCTTGCCATTTCCAGCCGGG-T	59
Db	113	CAGGCGCCACATGATGCTCCCGCCAAAGCTCCCGGGCCCGCTTGCCATTTCCAGCCGGGCT	172
QY	60	CCACCGAAGGTCACGCGCGCGGGAGAGGTGGAGCCCGAGAGCTCGCGCGGGGCCCG	115
Db	173	CCACCGAAGGTCACGCGCGCGGGAGAGGTGGAGCCCGAGAGCTCGCGCGGGGCCCG	233
QY	120	CCTGATGGCCGCGCCATGACAGCGGGTCGGAACTGGCTCTTTCCGGCGCCCTCCCG	172
Db	233	CCTGATGGCCGCGCCATGACAGCGGGTCGGAACTGGCTCTTTCCGGCGCCCTCCCG	292
QY	180	CGGAGGTGAGGGGAGATGTTCATGTCAGGGTTCAGAGCCCAACCGAAGTTACTGGCCCT	239
Db	293	CGGAGGTGAGGGGAGATGTTCATGTCAGGGTTCAGAGCCCAACCGAAGTTACTGGG-CT	351
QY	240	CTATCTTCCAGGAAACAGAGAGCCACGCGGGCTCACGCCCAACCGAATTAAG	299
Db	352	CTATCTTCCAGGAAACAGAGAGCCACGCGGGCTCACGCCCAACCGAATTAAGA	411
QY	300	TGA	302
Db	412	TTA	414

RESULT	3
LOCUS	HS2E10F
DEFINITION	HS2E10F 291 bp DNA PRI 18-OCT-1995
ACCESSION	H.sapiens Cpg island DNA genomic MseI fragment, clone 2e10, forward read cpg2e10.ft.1d.
NID	258126
VERSION	51029357
KEYWORDS	258126.1 GI:1029357
SOURCE	Cpg island; genomic MseI fragment.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 291)
AUTHORS	Macdonald,M., Hucklee,E., Wilkinson,P. and Micklem,G.
JOURNAL	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1HQ, England. E-mail contact: humquerry@sanger.ac.uk
REFERENCE	2 (bases 1 to 291)
AUTHORS	Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
JOURNAL	Purification of Cpg islands using a methylated DNA binding column Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE	94282070
COMMENT	Vector: pGEM-5zf(-) Clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1HQ, UK. See URL: http://www.hgmp.mrc.ac.uk/ for details or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES	1..291 Location/Qualifiers

BASE COUNT	47 a	99 c	102 g	48 t	2 others
ORIGIN					
BASE COUNT	44 a	95 c	103 g	45 t	4 others
ORIGIN					
Query Match	66.6%	Score 205.2	DB 9	Length 291	
Best Local Similarity	96.9%	Pred. No. 9.6e-34			
Matches 218	Conservative	0	Mismatches 6	Indels 1	Gaps 1
QY	1	CAGCGCCACATAGCTCCCGCCAGATCCTCGTGGCCCTTGCCATTTCAGCCGCG-T 59			
DB	67	CAGCGCCACATAGCTCCCGCCAGATCCTCGTGGCCCTTGCCATTTCAGCCGCGT 126			
QV	60	CCCACGAGGTTACGCGCGCGGAGAGGTGTGAAGCCGAGAGCTGCGCGGCGCCG 119			
	127	CCCACGAGGTTACGCGCGCGGAGAGGTGTGAAGCCGAGAGCTGCGCGGCGCGCCG 186			
QY	120	CCTGTTGCGCGCGCGCATACAGCGGCTCGGAGACAGCGCTTTCGCGCGCCCTCCGCG 179			
DB	187	CCTGTTGCGCGCGCGCATACAGCGGCTCGGAGACAGCGCTTTCGCGCGCCCTCCGCG 246			
QY	180	CGAGGTGAGGAGGAGATGTCCATGTCAGAGTTTCAAGCCCAAC 224			
DB	247	CGAGGTGAGGAGGAGATGTCCATGTCAGAGTTTCAAGCCCAAC 291			
RESULT	4				
HS2G7F					
LOCUS	HS2G7F	298 bp	DNA	PRI	19-OCT-1995
DEFINITION	H.sapiens Cpg island DNA genomic MseI fragment, clone 2g7, forward				
ACCESSION	Z60646				
NID	91032750				
VERSION	Z60646.1	GI:1032750			
KEYWORDS	Cpg island; genomic MseI fragment.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 298)				
JOURNAL	Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.				
COMMENT	Direct Submission				
REFERENCE	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire				
AUTHORS	CB10 IQO, England. E-mail contact: humquerry@sanger.ac.uk				
TITLE	2 (bases 1 to 298)				
JOURNAL	Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.				
COMMENT	Purification of Cpg islands using a methylated DNA binding column				
REFERENCE	Nat. Genet. 6 (3), 236-244 (1994)				
TITLE	Vector: pGEM-5zf(-)				
JOURNAL	Resources are available from the UK MRC Human Genome Mapping Project				
COMMENT	Clones are available from the UK MRC Human Genome Mapping Project				
REFERENCE	http://www.hgmp.mrc.ac.uk/ for details				
TITLE	or contact: bihelp@hgmp.mrc.ac.uk.				
JOURNAL	Location/Qualifiers				
COMMENT	1. 298				
FEATURES	/organism="Homo sapiens"				
SOURCE	/db_xref="taxon:9606"				
	/sex="male"				
	/dev_stage="adult"				
	/tissue_type="blood"				
	/clone_lib="CGI-1"				
	/clone="2g7"				

Query Match	58.8%	Score 181.2	DB 9	Length 288
Best Local Similarity	94.0%	Pred. No. 8.7e-29		
Matches 221	Conservative 0	Mismatches 9	Indels 5	Gaps 3
Qy 4	GCCTGACATAGCTCCGCCAAGTCTCTGTCGCCCTTCCCATTTTCCAGCCGC-TC	62		
Db 67	GCCTGACATAGCTCTCCGCCAAGTCTCTGTCGCCCTTCCCATTTTCCAGCCGC	126		
Qy 63	ACGAGGTCACGCGCGGGGAGAGTGAAGCCGGAAGCTCGGCCGGGGCCCGCT	122		
Db 127	ACGAGGTCACGCGCGGGGAGAGTGAAGCCGGAAGCTCGGCCGGGGCCCGCT	183		
Qy 123	GGTGGCCCGCGGCATGACAGAGCGGCTCGGGAGTGGCTCTTTTCGCGGCCCTCC	182		
Db 184	GGTGGCTCGCGGCACGACAGCGGCTCGGGAGAGGCTCTTTTCGCGGCCCTCC	243		
Qy 183	AGCTAGAGGGAAGATGTCCATGTACAGGTTTCAAGCCAAACGAAGTACTG	237		
Db 244	A-GTAGAGGGAAGATGTCCATGTCTCGGCTTCAAGCCAAACGAAGTACTG	297		
RESULT 5	AC005630	137699 bp	DNA	PRI 15-JAN-1999
LOCUS	AC005630			
DEFINITION	Homo sapiens clone DJ1129D05, complete sequence.			
ACCESSION	AC005630			
NID	94159882			
VERSION	AC005630.1	GI:4159882		
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 137699)			
REFERENCE	Waterston, R.H.			
JOURNAL	Submitted (08-SEP-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
COMMENT	On Jan 15, 1999 this sequence version replaced gi:3550965.			
FEATURES	Location/Qualifiers			
source	1..137699			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="DJ1129D05"			
BASE COUNT	36908 a 30670 c 32412 g 37709 t			
ORIGIN				
Query Match	58.8%	Score 181.2	DB 11	Length 137699
Best Local Similarity	80.3%	Pred. No. 3e-29		
Matches 249	Conservative 0	Mismatches 58	Indels 3	Gaps 3
Qy 1	CAGGGCGGCACATAGCT-CCGCGCAAGTCTCGGGCCCTTCCCATTTTCCAGCCGC	59		
Db 47458	CAGGGCGGCACATAGCTCTCCCGAGAGGCTCGGGTCCCTTCCCATTTTCCAGCCCTC	47517		
Qy 59	TCCACAGAGAGTACGCGCGGAGAGAGTGAAGCCGCGAGAGTCTCGCGCGGGGCGCC	118		
Db 47518	TCCGATTAAGTGAAGCATCAAGAGAGAGCGGAGAGTGGAGAGAGCGCGCGAGAGGTCT	47577		
Qy 119	GCCTGTCGCGCGGCATGACAGCGGCTCGGAGACTGCTCTTTTCGCGGCCCTTCCG	178		


```

ACCESSION      261392          read cpq5id8.rfa.
NID            91033770
VERSION        261392.1   GI:1033770
KEYWORDS       Cpg island; genomic MseI fragment.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE      Primates; Catarrhini; Homnidae; Homo.
AUTHORS        1 (bases 1 to 189)
TITLE          Macdonald,M., Huckie,E., Wilkinson,P. and Micklem,G.
JOURNAL        Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1RO, England. E-mail contact: humqueyes@sanger.ac.uk
REFERENCE      2 (bases 1 to 189)
AUTHORS        Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE          Purification of CpG islands using a methylated DNA binding column
JOURNAL        Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE        94282070
MENT           Vector: pGEN-5zf(-)
               clones are available from the UK MRC Human Genome Mapping Project
               Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See URL:
               http://www.hgmp.mrc.ac.uk/for details
               or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES       location/Qualifiers
source         1..189
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /sex="male"
              /dev_stage="adult"
              /tissue_type="blood"
              /clone_1lb="CGF-I"
              /clone="51d8"
BASE COUNT     32 a             61 c             71 g             22 t             3 others
ORIGIN
Query Match    16.0%; Score 49.4; DB 9; Length 189;
Best Local Similarity 84.8% ; Pred. No. 0.16;
Matches         Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 215 AGGCCAACCAGTACTTGCCCTCATCTTCACGGAGAACAAGAGCAGCACAGCGCGG 274
DB 1 A GGCAAGCTTAAGTTGTGG-CGCTACCTTCACGAACAGGAGCCGGCGCGG 59
OY 275 CTCACGCCCCACGCACACA 293
                ||||| | ||||| 
                60 CTCACGCTCCATGTCACACA 78

RESULT 13
LOCUS AC004690/c 142123 bp DNA PRI 16-MAY-1998
DEFINITION Homo saplens PAC clone DJ063OC24 from 7q31-q32, complete sequence.
ACCESSION AC004690
NID 93135287
VERSION AC004690.1 GI:3135287
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 142123)
TITLE Kalicki,J. and Burkhardt,J.
JOURNAL Unpublished (1998)
REFERENCE Unpublished (1998)
AUTHORS 2 (bases 1 to 142123)
WATERSTON,R. Direct Submission
SUBMITTED (16-MAY-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY: WUGSC
Genome Sequencing Center
```

Department of Genetics
Washington University
St. Louis MO 63108, USA
<http://genome.wustl.edu/gsc>
<mailto:saplens@watsn.wustl.edu>

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

The sequencing of this clone was established as part of a mapping and sequencing collaboration between the NGR1 Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7> or send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION: This clone was derived from human PAC library RPC1-4, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., *Nature Genetics* 6:84-9 (1994). The library is from one male donor. For further details, see <http://pacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of DJ0630C24;
actual end is at 142123 of DJ0630C24.

This clone contains STS SWS53567 (NID:q1916499).

This clone contains a transposon, from base 123435-124772, which has been deleted from the submitted sequence.

```

FEATURES
source
location/Qualifiers
1..142123
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="q31-q32"
/clone_l1b="RPC1-4"
/clone="D70630C24"
1..95
/rpt_family="Alu"
1135..1314
/rpt_family="MIR"
2378..2865
/rpt_family="MER50"
2964..3027
/rpt_family="MER50"
3463..3660
/rpt_family="Alu"
3735..3765
/rpt_family="L1"
3766..3858
/rpt_family="L1"
4018..4551
/rpt_family="L1"
7115..7409
/rpt_family="MER2_type"
8426..8775
/rpt_family="MIR"

```



```

repeat_region      9119. .9393
                    /rpt_family="Alu"
repeat_region      10494. .10794
                    /rpt_family="Alu"
repeat_region      11026. .11105
                    /rpt_family="MIR"
repeat_region      12496. .12797
                    /rpt_family="Alu"
repeat_region      13166. .13465
                    /rpt_family="Alu"
repeat_region      13480. .13797
                    /rpt_family="Alu"
repeat_region      14423. .15063
                    /rpt_family="MER21_g"
misc_feature        17610. .19237
                    /note="CpG_island (WGC=72.5, o/e=0.70, #CpGs=180)"
misc_feature        17778. .18095
                    /note="match to EST AA065302 (NID:g1929302)"
misc_feature        18792. .18846
                    /rpt_family="(CAG)n"
repeat_region      21367. .21491
                    /rpt_family="L2"
repeat_region      22129. .22260
                    /rpt_family="MIR"
repeat_region      22564. .22863
                    /rpt_family="Alu"
misc_feature        complement(23048. .23387)
                    /note="match to EST N79872 (NID:g1242573) y284c08.r1"
repeat_region      24581. .24871
                    /rpt_family="Alu"
repeat_region      26401. .26887
                    /rpt_family="L1"
repeat_region      27866. .28073
                    /rpt_family="L1"
repeat_region      28105. .30293
                    /rpt_family="Retroviral"
repeat_region      30452. .31613
                    /rpt_family="Retroviral"
repeat_region      31743. .32105
                    /rpt_family="MALR"
repeat_region      32247. .32452
                    /rpt_family="L2"
repeat_region      32540. .32702
                    /rpt_family="MIR"
repeat_region      32990. .33364
                    /rpt_family="MALR"
repeat_region      34851. .35026
                    /rpt_family="MER1_type"
repeat_region      35047. .35314
                    /rpt_family="MALR"
repeat_region      35315. .35544
                    /rpt_family="MALR"
repeat_region      35556. .35682
                    /rpt_family="MALR"
repeat_region      35856. .36160
                    /rpt_family="Alu"
repeat_region      36809. .37273
                    /rpt_family="MALR"
repeat_region      38194. .38495
                    /rpt_family="Alu"
repeat_region      38611. .38908
                    /rpt_family="Alu"
repeat_region      41068. .41370
                    /rpt_family="Alu"
repeat_region      41514. .41938
                    /rpt_family="MALR"
repeat_region      42534. .42982
                    /rpt_family="MALR"
repeat_region      45202. .45318
                    /rpt_family="Retroviral"
repeat_region      45343. .45395
                    /rpt_family="MALR"
repeat_region      45461. .45624
                    /rpt_family="MIR"

```

```

repeat_region      /rpt_family="Alu"
                    45759. .45881
repeat_region      /rpt_family="L1"
                    45902. .46280
repeat_region      /rpt_family="L1"
                    48272. .48501
repeat_region      /rpt_family="L1"
                    48505. .49036
repeat_region      /rpt_family="L1"
                    49052. .49652
repeat_region      /rpt_family="L1"
                    49647. .50340
repeat_region      /rpt_family="MALR"
                    50341. .50720
repeat_region      /rpt_family="MALR"
                    50723. .51658
repeat_region      /rpt_family="L1"
                    52203. .52314
repeat_region      /rpt_family="L2"
                    52505. .52616
repeat_region      /rpt_family="MER1_type"
                    52617. .52789
repeat_region      /rpt_family="Alu"
                    52790. .52873
repeat_region      /rpt_family="MER1_type"
                    52937. .53912
repeat_region      /rpt_family="L2"
                    54049. .54356
repeat_region      /rpt_family="Alu"
                    54511. .55100
repeat_region      /rpt_family="Retroviral"
                    56013. .56227
repeat_region      /rpt_family="MIR"

```

Query Match 14.6%; Score 45; DB 11; Length 142123;
 Best Local Similarity 53.8%; Pred. No. 0.42; Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```

QY 18 CCGGCCAGTCCTCGTCCGTCCTTCCATTTCCAGCCGCTCCACAGAGGTACAGGCG 77
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 18251 CCGGCCATGGGCTCGAGGTGTCCGACGCCCGAGGCTGCACCGCGCTGCCGCGGT 18192

QY 78 GCGGGAGAGTGTGACGCCGAGAGCTCGGCGGGGCCCCGCTGTGGCCCGGCCCA 137
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 18191 GCGGGAGTGTGACGCCGAGGCGCCGAGGCCCGGCCCGCTCTGTGCGCGTCCGCCC 18132

QY 138 GACAGCGCTCGGAGTGTCTCTTTCCGCGCCCTCCCGCCGAGGTGAGG 190
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 18131 CCGGGGGGCGGAGGAGCTCGTCCGCGCGGTCCCGGCGGAGCGGGG 18079

```

RESULT 14
 HSCG16P5/c 2837 bp DNA 13-NOV-1998
 LOCUS Home sapiens DNA for (CGG)n trinucleotide repeat region, isolate
 DEFINITION
 ACCESSION AJ001218
 NID 93780170
 VERSION AJ001218.1 GI:3790170
 KEYWORDS trinucleotide repeat.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 2837)
 AUTHORS Mangel,L.
 TITLE Direct Submission
 JOURNAL Submitted (12-AUG-1997) Mangel L., Medical Genetics and Virology,
 University of Cologne, Institute of Genetics, Weyertal 121,
 Cologne, 50931, GERMANY
 REMARK Revised by [3]
 REFERENCE 2 (bases 1 to 2837)
 AUTHORS Mangel,L.

Sun Sep 19 10:27:09 1999

us-08-956-518a-95.rge

Page 10

Db 3633 GCACCGCTCGAGCGCTCGACCGGTCTCGCCGGGGGGGTGGGGCCGAGCCGCGCTAC 3692
OY 233 TGGCCCTTATCTTCCAGAGAACCGAGGAGCCACAGCCGGGCTCAGGCCCCACCGCA 290
Db 3693 CGCGACCCGTCGGGGGGCGAGCTGTGGGCTCGACGTGGCCAGGCGCAGCAGCA 3750

Search completed: September 17, 1999, 22:03:03
Job time: 16366 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:10 ; Search time 2825.05 seconds
(without alignments)
215.055 Million cell updates/sec

Title: US-08-956-518a-95

Perfect score: 308
Sequence: 1 CAGGCCGCCACATAGCTCC.....CAACATTAGGTAGTCGCC 308

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

base :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est23: *
55: em_est24: *
56: em_est25: *
57: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	93.2	30.3	170	25	N83152	N83152 K4329F Huma
2	81.2	26.4	510	54	HSN011948	A1047098 Homo Sap1
3	57.6	18.7	335	32	AA333244	AA333244 EST03117
4	43.8	14.2	416	33	AA407437	AA407437 EST00734
5	42	13.6	328	33	AA407004	AA407004 EST02243
6	41.6	13.5	213	33	AA410084	AA410084 EST01971
7	40.4	13.1	268	33	AA407006	AA407006 EST02245
8	40.4	13.1	229	33	AA410099	AA410099 EST01951
9	40.4	13.1	325	44	A1317247	A1317247 u136c11.Y
10	40	13.0	371	47	A1508264	A1508264 mb6603.Y
11	39.8	12.9	374	33	AA407075	AA407075 EST02080
12	39.8	12.9	376	33	AA407297	AA407297 EST02294
13	39.8	12.9	271	33	AA407662	AA407662 EST01743
14	39.8	12.9	316	33	AA407690	AA407690 EST01973
15	39.8	12.9	373	33	AA407899	AA407899 EST01001
16	39.8	12.9	322	33	AA407912	AA407912 EST01757
17	39.8	12.9	224	33	AA407943	AA407943 EST01023
18	39.8	12.9	335	33	AA408014	AA408014 EST02534
19	39.8	12.9	388	33	AA408056	AA408056 EST01037
20	39.8	12.9	354	33	AA408063	AA408063 EST01129
21	39.8	12.9	376	33	AA408103	AA408103 EST02583
22	39.8	12.9	321	33	AA408119	AA408119 EST02666
23	39.8	12.9	261	33	AA408198	AA408198 EST02863
24	39.8	12.9	321	33	AA408495	AA408495 EST02965
25	39.8	12.9	437	33	AA408575	AA408575 EST03109
26	39.8	12.9	485	33	AA408661	AA408661 EST03176
27	39.8	12.9	414	33	AA408707	AA408707 EST03168
28	39.8	12.9	525	33	AA408911	AA408911 EST03425
29	39.8	12.9	308	33	AA409494	AA409494 EST01256
30	39.8	12.9	315	33	AA410085	AA410085 EST01972
31	39.8	12.9	372	45	A1324211	A1324211 m122c07.X
32	39.8	12.9	334	45	A1324741	A1324741 mb78c09.X
33	39.8	12.9	373	45	A1324799	A1324799 mc31n01.X
34	39.8	12.9	356	46	A1414218	A1414218 md61e11.X
35	39.8	12.9	365	46	A1414287	A1414287 md81f02.X
36	39.8	12.9	372	46	A1426055	A1426055 mg53a11.X
37	39.8	12.9	335	46	A1430785	A1430785 mb66003.X
38	39.8	12.9	372	46	A1449591	A1449591 mq06e08.X
39	39.8	12.9	354	46	A1452165	A1452165 mp81e11.X
40	39.8	12.9	372	47	A1463228	A1463228 uc39p02.X
41	39.8	12.9	372	50	A1666386	A1666386 ml14a08.X
42	38.6	12.5	369	29	AA156428	AA156428 z151c02.S
43	38.4	12.5	312	33	AA10147	AA10147 EST01892
44	38.4	12.5	372	49	A1645460	A1645460 v188b01.X
45	38.2	12.4	354	46	A1426052	A1426052 mg52e10.X

ALIGNMENTS

RESULT 1
N83152
LOCUS N83152 170 bp mRNA
DEFINITION K4329F Human fetal heart, Lambda ZAP Express Homo sapiens CDNA
ACCESSION N83152
NID g1441866
VERSION N83152.1 GI:1441866

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Liew,C.C.
JOURNAL CDNA from fetal heart (1996)
COMMENT Unpublished (1996)
On Apr 14, 1993 this sequence version replaced g1:502850.

Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Seq primer: GAATTAACCTCCTCAAGG.
Location/Qualifiers
1. 170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="K4329"
/lab_host="E. coli XL1-Blue"
/lab="lab-1b="Human fetal heart, Lambda ZAP Express"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). CDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."

BASE COUNT 43 a 47 c 46 g 34 t
ORIGIN

Query Match 30.3%; Score 93.2; DB 25; Length 170;
Best Local Similarity 83.1%; Pred. No. 1.1e-11;
Matches 118; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

QY 155 GGGTCCTTTCCGGCCCGCCGCGAGAGTGAAGGAAAGATGTCAGGGTTCA 214
DB 9 GGGTCCTTTCCGGCCCGCCGCGAGATGAGGGAAGATGTGTCAAGATTCA 68
QY 215 AGCCCAACGAGATGAGGCTGATCTCCAGAGAACGAGACGACGCGCG 274
DB 69 AGCCCAACGAGATGAGGCTGATCTCCAGAGAACGAGAGGCTCAACTGCGG 127
275 CTCAGCGCCCGACGCAACATTA 296
DB 128 CTCAGGCTCCTCAACATGA 149

RESULT 2

HS011948 standard; RNA; EST; 510 BP.

AC AL047098;
SV AL047098.1
NT e1405574
XX
XX
DE 12-MAR-1999 (Rel. 59, Created)
XX 12-MAR-1999 (Rel. 59, Last updated, Version 1)
DE Homo sapiens mRNA; EST DKFZp586B2218_r1 (from clone
DE DKFZp586B2218)
XX
XX EST; expressed sequence tag.
XX Homo sapiens (human)
XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
XX

OC Primates; Catarrhini; Homnidae; Homo.
XX
XX [1]
RN 1-510
RP Koehler K., Beyer A., Mewes W., Gassenhuber J., Wiemann S.;
RA Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RT MIPs, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
RL
CC Clone from S. Wiemann, sequenced by BMFZ within the CDNA
CC sequencing consortium of the German Genome Project
CC No s1 sequence available
CC This clone is available at the RZPD in Berlin
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX
FH Key Location/Qualifiers
FT 1. 510
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKFZp586B2218"
FT /clone_1lb="586 (synonym: hutel). Vector pSport1; host
FT DH10B; sites NotI + SalI/MluI"
FT /dev_stage="adult"
FT /tissue_type="uterus"

Sequence 510 BP; 118 A; 154 C; 150 G; 88 T; 0 other;

Query Match 26.4%; Score 81.2; DB 54; Length 510;
Best Local Similarity 73.3%; Pred. No. 5.4e-09;
Matches 154; Conservative 0; Mismatches 33; Indels 23; Gaps 3;

QY 78 GCGGGAGAGTGTGAGCCGCGAGAGTGGCC---GGGGCCCCCGCTGTGCGCGGC 134
DB 3 GCTCGAGAGGCGGAGAGCCCGAGAGCGCGCCAGCGCCCGGGCGGTGCGCGC 62
QY 135 CATGACAGCGGCTCGGACT-----GGTCCTTTCCGGCCCGCC 175
DB 63 CGTGAAGGCGGCTCGGGCGCGCTCCCTTCGCGCGCGCTCCCTTCGCGCGC 122
QY 176 CCGCGGAGGTGAGGGAAGATGTCATGTCAGGCTCAAGGCCAAGAGTACTGG 235
DB 123 CCGCGGAGATGAGGGAAGATGTCGATGTCAGGCTCAAGGCCGAGTGAATTCCTGG 182
QY 236 CCTCTATCTTCCAGAGAACGAGAGCCA 265
DB 183 -GTCATCTTCCAGACAGACGAGCGA 211

RESULT 3

AA333244 335 bp mRNA EST 21-APR-1997
LOCUS AA333244
DEFINITION EST37317 Embryo, 8 week I Homo sapiens CDNA 5' end similar to
similar to interleukin-2 receptor, alpha chain, kappa B binding
protein, mRNA sequence.

ACCESSION AA333244
NID 01985508
VERSION AA333244.1 GI:1985508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 335)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.D., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Gloddek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,D.M., Kelley,D.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligriano, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shiley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Danks, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

On May 18, 1995 this sequence version replaced g1:811121.

96026280

JOURNAL MEDLINE

COMMENT

Other ESTs: THC185696

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..335

/organism="Homo sapiens"

/db_xref="ATCC (Inhost):134920"

/db_xref="taxon:9606"

/clone_lib="Embryo, 8 week 1"

/dev_stage="Embryo, 8 wks"

/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-"

Site_1: EORI; Site_2: XhoI"

84 a 94 c 88 g 69 t

BASE COUNT

ORIGIN

Query Match 18.7%; Score 57.6; DB 32; Length 335;

Best Local Similarity 83.7%; Pred. No. 0.00069;

Matches 77; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 174 TCCGCGCGGAGTGGAGGAGATGTCATGTCAGGCTCAAGCCAAACGAGTACT 233

1 TCCCGCGGAGATGAGGAGGAGATGTCGTCGTACAGGCTCAAGCCGAGCTGAATTCT 60

QY 234 GGCCTCTATCTTCCAGAGACCAAGAGCCA 265

DB 61 GG-CGTCCATCTTCAGACAGACCAAGAGCCA 91

RESULT 4

AA407437 416 bp mRNA EST 26-AUG-1998

LOCUS AA407437

DEFINITION EST00734 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone C0002B08 3', mRNA sequence.

ACCESSION AA407437

NID 92065649

VERSION AA407437.1 GI:2065649

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 416)

Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cui, Y., Wang, X., Pryor, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsunoto, S. and Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation site

JOURNAL

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced g1:1292065.

Contact: KO MSH

Center for Molecular Medicine and Genetics

Wayne State University

5047 Guillen Mall, Detroit, MI 48202

Tel: 3135776708

Fax: 3135776200

Email: makoeemb@iosci.wayne.edu

Human 28S ribosomal RNA gene.

Seq primer: M13 Forward.

Location/Qualifiers

1..416

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="ATCC (Inhost):1363489"

/db_xref="taxon:10090"

/map="20"

/clone="C0002B08"

/clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"

/sex="unknown"

/tissue_type="ectoplacental cone"

/dev_stage="embryonic day 7.5 postconception"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

67 a 130 c 133 g 69 t 17 others

BASE COUNT

ORIGIN

Query Match 14.2%; Score 43.8; DB 33; Length 416;

Best Local Similarity 54.2%; Pred. No. 0.73;

Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 28 CCTCGTGCCCTTTCATTTCCAGCCGCTCCAGAGGTCACGGCGGCGGAGAG 87

DB 41 CCGCGGGCGCNCNGCNCACCCCGCCGCTCTCCGAGGGCGGCGGCGCGC 100

QY 88 GTGAGCCGCGAGAGCTCGCGCGGCGCGCGCTGTGTGTCGCGCCGCTGACAGCGGT 147

DB 101 GCGCTCCCGCGGTGTGTGTGAGACCTCCGCTGTGTCCCGCGCGGTCCGCC 160

QY 148 CGGAGCTGCTCTTTCCGCGCCCTCCGCGCG 182

DB 161 CGGCGCGCGTTTCCGCGCGCGCGCTCGCTCGG 195

RESULT 5

AA407004 328 bp mRNA EST 26-AUG-1998

LOCUS AA407004

DEFINITION EST02243 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone C0019B05 3', mRNA sequence.

ACCESSION AA407004

NID 92067158

VERSION AA407004.1 GI:2067158

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 328)

Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cui, Y., Wang, X., Pryor, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsunoto, S. and Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation site

JOURNAL

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced g1:1393284.

FEATURES
SOURCE

BASE COUNT ORIGIN	9 a	88 c	81 g	28 t	7 others
----------------------	-----	------	------	------	----------

[illegible]

source

BASE COUNT	9 a	88 c	81 g	28 t	7 others
ORIGIN					

	Query Match	13.5%;	Score 41.6;	DB 33;	Length 213;
	Best Local Similarity	52.5%;	Pred. No. 1.9;		
Matches	83; Conservative	07; Mismatches	75; Indels	0; Gaps	0;
QY	28 CCTCGTGCCCCCATTGCCATTTTCCAGCCGSCGTCCCACAGAGGGTCAGGGCGGGGAGAG	87			
Dδ	42 CC GG GGGGG CCT CG CTGC ACC CCCCA AT GC CTC TCC AGA GT GC TT GCC TG CG GGG GGG GGG	101			
QY	88 GTGAGCGCGGAGAAGCTCCGCGGGGGCCCGCTGGTGCGCGCGCATGACAGCGCT	147			
Dβ	102 GCGCTCCCGCGCGGTGGGGGAACACTCCGCGGTGGGTATTTCCCGCCGCGGTGACNCCCCC	161			
QY	148 CGGAGACTGACTCTTTTCCGCGGCCCTCCGCGGAGG	185			
Dγ	162 CGGCGCGNGTTTTCCGCGGGGCCCTINGCNTGCGGCMG	199			

RESULT	7
LOCUS	AA407006
DEFINITION	AA407006 268 bp mRNA EST 26-AUG-1998
ACCESSION	AA407006
VERSION	92067160
KEYWORDS	AA407006.1 GI:2067160
SOURCE	EST.
ORGANISM	house mouse. house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eumammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Muridae; Mus

TITLE	Systematic analyses of mouse genes expressed in embryo implantation site
JOURNAL	Unpublished (1997)
COMMENT	On Sep 12, 1996 this sequence version replaced g1.1393286.

Contact: KO MSH

Contact: KO MSH
Center for Molecular Medicine and Genetics
Wayne State University

JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1393842

Contact: Ko MSH
Center for Molecular Medicine and Genetics
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: mskocemc.biosci.wayne.edu
human 28S ribosomal RNA gene.
Seq primer: M13 forward.

BASE COUNT	55 a	126 c	132 g	61 t
ORIGIN				

[illegible]

RESULT	12
LOCUS	AA407297
DEFINITION	AA407297 376 bp mRNA EST02294 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone C0019501 3', mRNA sequence.
ACCESSION	AA407297
NID	92067209
VERSION	AA407297.1
KEYWORDS	GI:2067209
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 376) Ko,M.S.H., Threat,T.A., Horton,J.H., Wang,X., Cui,Y., Wang,X., Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and Nakashima,H.
TITLE	Systematic analyses of mouse genes expressed in embryo implantation site
JOURNAL	Unpublished (1997)
COMMENT	On Sep 12, 1996 this sequence version replaced gi 1393086.

CONTACT: Ko MSH
 Center for Molecular Medicine and Genetics
 Wayne State University
 5047 Gullen Mall, Detroit, MI 48202
 Tel: 3135776708
 Fax: 3135776200
 Email: mskocemb.biosci.wayne.edu
 Human 28S ribosomal RNA gene.
 Seq primer: M13 Forward.
LOCATION/QUALIFIERS
 1. .376

BASE COUNT	57 a	.126 c	132 g	61 t
ORIGIN				

[illegible]

RESULT	13			
LOCUS	AA407662			
DEFINITION	EST01743 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus			
ACCESSION	AA407662	271 bp	mRNA	EST
NID	AA407662			
VERSION	q2066658			
KEYWORDS	AA407662.1	GI:2066658		
SOURCE	EST.			
ORGANISM	house mouse.			
	mus musculus			
REFERENCE	Eukaryote; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;			
AUTHORS	Euthelia; Rodentia; Scutognath; Muridae; Murine; Mus.			
	1 (bases 1 to 271)			
	Ko,M.S.H., Theact,T.A., Horton,J.H., Wang,X, Cui,Y., Wang,X.,			
	Plyor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and			
	Nakashima,H.			
TITLE	Systematic analyses of mouse genes expressed in embryo implantation			
	site			
JOURNAL	Unpublished (1997)			
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1393367.			
CONTACT:	Ko MSH			

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:03 ; Search time 1962.17 Seconds
(Without alignments)
235.018 Million cell updates/sec

Title: US-08-956-518A-96
Perfect score: 145
Sequence: 1 CTCATTCAGATTACAAAGTG.....ACTCTGACAGCTGAGTTGTA 145

Scoring table:

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank:
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sy:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_cm:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_st:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129.2	89.1	641	11	AF029838 Homo sapi

2	128.8	88.8	689	11	AF029839	AF029839 Homo sapi
3	34.4	23.7	3093	12	RND0CP	X13417 Rat ornithi
4	30.4	21.0	491	7	MINC01	V00667 Neurospora
5	30.2	20.8	143749	11	AC004962	AC004962 Homo sapi
6	30.0	20.7	146058	11	AC003668	AC003668 Homo sapi
7	29.8	20.6	196589	34	AF128834	AF128834 Homo sapi
8	29.8	20.6	132050	42	AC004909	AC004909 Homo sapi
9	29.4	20.3	2860	8	AF100163	AF100163 Arabidops
10	29.4	20.3	105654	8	ATAF001308	AF001308 Arabidops
11	29.4	20.3	73289	19	AC005981	AC005981 *** SEQUE
12	29.4	20.3	81146	35	AC005427	AC005427 Drosophi
13	29.4	20.3	62484	35	AC005453	AC005453 Drosophi
14	29.4	20.3	35807	36	CET13H5	Z6524 Caenorhabd
15	29.4	20.3	60019	37	AC005650	AC005650 Drosophi
16	29.2	20.1	100639	34	AC002417	AC002417 Homo sapi
17	28.6	19.7	2036	3	P1GSLTXX	L02900 Pig Na-depe
18	28.6	19.7	2542	7	ATC1CB	Z71446 A.thaliana
19	28.6	19.7	2755	8	AF133209	AF133209 Nicotiana
20	28.6	19.7	1354	10	HSN800075	AL050297 Homo sapi
21	28.4	19.6	12631	2	AE001282	AE001282 Chlamydia
22	28.4	19.6	175120	11	AC004687	AC004687 Homo sapi
23	28.4	19.6	6309	12	AF015311	AF015311 Rattus no
24	28.4	19.6	173616	35	AC006509	AC006509 Homo sapi
25	28	19.3	4169	7	ANSAMB	AJ000996 Aspergill
26	28	19.3	2040	7	OSCDPK2	X81394 O.saliva MR
27	27.8	19.3	102402	35	AC007641	AC007641 Mus muscu
28	27.8	19.2	100000	9	AP000068	AP000068 Homo sapi
29	27.8	19.2	165139	10	HS436K10	AL031407 Human DNA
30	27.8	19.2	273800	12	AF100956	AF100956 Mus muscu
31	27.8	19.2	213245	12	MMHC425018	AF110520 Mus muscu
32	27.8	19.2	112726	34	CET21E11	AL009107 Caenorhab
33	27.8	19.2	281947	35	AC007569	AC007569 Homo sapi
34	27.8	19.2	94718	36	AC004658	AC004658 Drosophi
35	27.8	19.2	26208	36	CCE40C9	Z70266 Caenorhabd
36	27.8	19.2	3722	36	CER42H11	Z83225 Caenorhabd
37	27.6	19.0	3781	8	AF030634	AF030634 Neurospor
38	27.4	18.9	1541	8	ATU80185	U80185 Arabidopsis
39	27.4	18.9	100000	9	AP000147	AP000147 Homo sapi
40	27.4	18.9	89086	9	AP000233	AP000233 Homo sapi
41	27.4	18.9	3340	12	AF136401	AF136401 Rattus no
42	27.4	18.9	171914	34	AC006169	AC006169 Drosophi
43	27.4	18.9	85096	35	AC007323	AC007323 Arabidops
44	27.2	18.8	7150	9	AB007931	AB007931 Homo sapi
45	27.2	18.8	36133	9	HSA007973	AB007973 Homo sapi

ALIGNMENTS

RESULT 1	AF029838	641 bp	MRNA	PRI	16-DEC-1998
LOCUS	Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.				
DEFINITION	AF029838				
ACCESSION	AF029838				
NID	93757793				
VERSION	AF029838.1	GI:3757793			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
AUTHORS	Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Buiting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.				
TITLE	Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene				
JOURNAL	Genomics 52 (2), 173-185 (1998)				
MEDLINE	99000837				
REFERENCE	2 (bases 1 to 641)				
AUTHORS	Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.				
TITLE	Direct Submission				

JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES
Source
1. 641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR1"
/note="RACE product A/B/C/D"
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

misc_feature
151 a 192 c 163 g 135 t

BASE COUNT 151 a 192 c 163 g 135 t

ORIGIN

Query Match 89.1%; Score 129.2; DB 11; Length 641;
Best Local Similarity 97.8%; Pred. No. 7.1e-33;
Matches 131; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 TTTGATTACAGTGTGACACCTGAGTACAGCAGCTGGAATCCAGATGAGAGCTTA 65
|||
Db 294 TTAAGATTACAGTGTGACACCTGAGTACAGCAGCTGGAATCCAGATGAGAGCTTA 353

QY 66 TCTACAGACTGAGATCTGTGTGTACACCCCATTTATGACAAATCCAAAGGTGCAGAAAGC 135
|||||
Db 354 TCTACAGACTGAGATCTGTGTGTACACCCCATTTATGACAAATCCAAAGGTGCAGAAAGC 413

QY 126 ACTCTGACAGTGA 139
|||||
Db 414 ACTCTGACAAATGA 427

RESULT 2
AF029839 689 bp mRNA PRI 16-DEC-1998
LOCUS AF029839
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
ACCESSION AF029839
NID 93757794
VERSION AF029839.1 GI:3757794
KEYWORDS
SOURCE human:
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 689)
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Buiting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL 99000837
MEDLINE 2 (bases 1 to 689)
REFERENCE
AUTHORS Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES
Source
1. 689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

misc_feature
142 a 210 c 196 g 141 t

BASE COUNT 142 a 210 c 196 g 141 t

ORIGIN

Query Match 88.8%; Score 128.8; DB 11; Length 689;
Best Local Similarity 98.5%; Pred. No. 9.7e-33;
Matches 130; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TTTGATTACAGTGTGACACCTGAGTACAGCAGCTGGAATCCAGATGAGAGCTTA 65
|||
Db 406 TTAAGATTACAGTGTGACACCTGAGTACAGCAGCTGGAATCCAGATGAGAGCTTA 465

QY 66 TCTACAGACTGAGATCTGTGTGTACACCCCATTTATGACAAATCCAAAGGTGCAGAAAGC 125
|||||
Db 466 TCTACAGACTGAGATCTGTGTGTACACCCCATTTATGACAAATCCAAAGGTGCAGAAAGC 525

QY 126 ACTCTGACAGT 137
|||||
Db 526 ACTCTGACAAAT 537

RESULT 3
R000CP/c 3093 bp DNA ROD 06-JUL-1989
LOCUS R000CP/c
DEFINITION Rat ornithine decarboxylase pseudogene.
ACCESSION X13417
NID 956788
VERSION X13417.1 GI:56788
KEYWORDS ornithine decarboxylase; pseudogene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 3093)
Oka, T.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1988) to the EMBL/Genbank/DBJ databases
AUTHORS Kanamoto, R., Yoshimura, M., Hayashi, S. and Oka, T.
REFERENCE
1 (bases 1 to 3093)
Nucleotide sequence of a pseudogene for rat ornithine decarboxylase
JOURNAL Nucleic Acids Res. 17 (1), 463 (1989)
MEDLINE 89098420
COMMENT Data kindly reviewed (20-Apr-1989) by Oka T.
FEATURES
Source
1. 3093
/organism="Rattus norvegicus"
/strain="Sprague Dawley"
/db_xref="taxon:10116"

misc_feature
615. 1618
/note="Ornithine decarboxylase pseudogene region"

BASE COUNT 857 a 683 c 623 g 930 t

ORIGIN

Query Match 23.7%; Score 34.4; DB 12; Length 3093;
Best Local Similarity 57.4%; Pred. No. 0.2;
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 CTCATTTCAGATTACAGAGGAGACCTGAGTACAGCAGCTGGAATCCAGATGAGAGC 60
|||
Db 1488 CTCCTTCACACACTTTCAGCGGCAATGATGTCAACGCAAGTGTGAAGTAGAGAGACA 1429

QY 61 GCTTATCTACAGCAGCTGAGATCTGTGTACACCCCATTTATGACAAAT 108
|||||
Db 1428 CAGTATCTCCCGGCTCAGTTATGTTTCACACCTCAGAGCTGATGAT 1381

RESULT 4
MINCO1/c 491 bp DNA PLN 21-OCT-1996
LOCUS MINCO1/c
DEFINITION Neurospora crassa ATPase proteolipid like gene and unidentified
ACCESSION V00667 J01428
NID 92938
VERSION V00667.1 GI:2938

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:16 ; Search time 2025.05 seconds

(without alignments)
58.651 Million cell updates/sec

Title: US-08-956-518a-97

Sequence: 1 TTAAACACAGATATATGAAAC.....GCATTTTCAGTAGATCAT 84

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

abase :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45.2	53.8	501	49	A1655990
2	38	45.2	500	26	W52861
3	31.6	37.6	487	39	AA861176
4	31.2	37.1	457	44	AI286133
5	28.8	34.3	549	34	AA475853
6	28.2	33.6	420	49	AT631518
7	28	33.3	567	49	AU056388
8	27.8	33.1	362	21	D47311
9	27.6	32.9	395	23	H47049
10	27.6	32.9	353	23	R93050
11	27.6	32.9	202	32	AA360272
12	27.6	32.9	430	32	AA398114
13	27.6	32.9	493	35	AA553441
14	27.4	32.6	496	35	AA572984
15	27.2	32.4	567	48	AU051905
16	27	32.1	557	49	AI622332
17	27	32.1	696	49	AU057033
18	26.6	31.7	563	49	AT651377
19	26.4	31.4	322	20	T40435
20	26.4	31.4	379	21	R09085
21	26.4	31.4	366	21	T80555
22	26.4	31.4	506	22	R66628
23	26.4	31.4	316	24	H98196
24	26.4	31.4	449	24	H99885
25	26.4	31.4	451	25	N55011
26	26.4	31.4	441	25	W12683
27	26.4	31.4	582	26	W54506
28	26.4	31.4	495	26	W62896
29	26.4	31.4	468	26	W77761
30	26.4	31.4	580	27	AA003956
31	26.4	31.4	512	27	AA010721
32	26.4	31.4	582	27	AA016534
33	26.4	31.4	619	27	AA030762
34	26.4	31.4	573	28	AA075433
35	26.4	31.4	661	28	AA124475
36	26.4	31.4	496	31	AA310074
37	26.4	31.4	464	31	AA314005
38	26.4	31.4	280	32	AA334911
39	26.4	31.4	344	32	AA340148
40	26.4	31.4	310	32	AA344762
41	26.4	31.4	199	32	AA373126
42	26.4	31.4	660	33	AA394081
43	26.4	31.4	556	33	AA405892
44	26.4	31.4	370	33	AA442497
45	26.4	31.4	826	50	AU067186

ALIGNMENTS

RESULT 1
LOCUS A1655990/c
DEFINITION tt42c03.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2243428 3',
ACCESSION A1655990
NID 94739969
VERSION A1655990.1 GI:4739969

EST 04-MAY-1999

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 501)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced g1:3188057.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bhrp/image/image.html

Seq primer: -400p from Glibco
High quality sequence stop: 458.
Location/Qualifiers
1. 501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3243428"
/clone_1lb="NCI-CGAP_GC6"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Plasmid DNA from the normalized library
NCI-CGAP_GC6 was prepared, and ss circles were made in
vivo. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clonoids 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 107 c 115 g 144 t

ORIGIN

Query Match 53.8%; Score 45.2; DB 49; Length 501;
Best Local Similarity 90.9%; Pred. No. 0.00066;
Matches 60; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 9 AGATATGAAACACCATCGTTAATTGATGCAAAATATTGATCTACACGAT 68
|||||
Db 501 AATATATGAAACACGACCA---GTTAATTGATGCAAAATATTGATCTACACAT 445
|||||

QY 63 TTTCAG 74
|||||

Db 444 ATTTCAG 439

RESULT 2
W52861 560 bp mRNA EST 10-OCT-1996
LOCUS zc03e02.r1 Soares-parathyroid_tumor_NBHPA Homo sapiens cDNA clone
IMAGE:321242.5, similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN), mRNA sequence.
ACCESSION W52861
NID G1350351
VERSION W52861.1 GI:1350351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 560)
TITLE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Jan 24, 1995 this sequence version replaced g1:634363.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1753 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 408.
Location/Qualifiers
1. 360
/organism="Homo sapiens"
/db_xref="GDB:1258900"
/db_xref="taxon:9606"
/clone="IMAGE:321242"
/clone_1lb="Soares-parathyroid_tumor_NBHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pRTT3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
15'-
TGTTCACATCTGATGAGTGGAGCGCCGACCAATTTTTTTTTTTTTTTT
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pRTT3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 128 a 137 c 137 g 152 t 6 others

ORIGIN

Query Match 45.2%; Score 38; DB 26; Length 560;
Best Local Similarity 89.1%; Pred. No. 0.076;
Matches 41; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 36 AATTGATGCAAAATATTGATCTACACGATTTTCAGTAGAT 81
|||||
Db 1 AATTGATGCAAAATATTGATCTACACGATTTTCAGTTCCAT 46
|||||

RESULT 3
AA861176 487 bp mRNA EST 04-JAN-1999
LOCUS AK36409.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1408049
DEFINITION 3', mRNA sequence.
ACCESSION AA861176
NID 92953316
VERSION AA861176.1 GI:29533316
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 487)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 9, 1998 this sequence version replaced gi:3937098.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/db/ftp/image/image.html

Insert Length: 758 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 458.
 Location/Qualifiers

1. 487
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:140804.9"
 /clone_11d="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5].
 TGTACCAATCGATGAGGAGGCGCCCAATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 131 a 95 c 102 g 159 t

Query Match 37.6%; Score 31.6; DB 39; Length 487;
 Best Local Similarity 89.5%; Pred. No. 5.2;
 Matches 34; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

44 GCAAAATATTCATCTACCAAGCATTTTTCAGTAGAT 81
 487 GCAAAATATTCATCTACCAAGCATTTTTCAGTTCCAT 450

RESULT 4
 LOCUS A1286133
 DEFINITION q101c05.x1 Soares.NFL.T.GBC.S1 Homo sapiens CDNA clone
 IMAGE:1855208 3' similar to gb:U74525 UBIQUITIN-CONJUGATING ENZYME
 E2-17 KD (HUMAN); mRNA sequence.
 ACCESSION A1286133
 NID 93924366
 VERSION A1286133.1 GI:3924366
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 457)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2044002.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (infoimage.lnl.gov) for further information.
 Insert Length: 787 Std Error: 0.00
 Seq primer: -40UP from Gldco.
 High quality sequence stop: 370.
 Location/Qualifiers

1. 457
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1855208"
 /clone_11d="Soares.NFL.T.GBC.S1"
 /lab_host="DH10B"
 /note="Organ: pooled. Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker. Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 128 a 89 c 94 g 144 t 2 others
 ORIGIN

Query Match 37.1%; Score 31.2; DB 44; Length 457;
 Best Local Similarity 63.2%; Pred. No. 6.8;
 Matches 48; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 9 AGATATGAACAACCAACCATCGTTAAATTTGATCAAAATATTCATCTACCAAGAT 68
 DB 399 AGATATGAACAACCAACCATCTAGATTGTCTCTAAGATGTCATCCAAATGCT 340
 QY 69 TTTCAGGTAGATCAT 84
 DB 339 ATCCAGATGCTAGTAT 324

RESULT 5
 LOCUS AA475853
 DEFINITION vhl1g04.r1 Soares mouse mammary gland NBMMG Mus musculus CDNA clone
 IMAGE:875286 5', mRNA sequence.
 ACCESSION AA475853
 NID 92203704
 VERSION AA475853.1 GI:2203704
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Nov 29, 1993 this sequence version replaced gi:430168.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

RESULT 5
H47049

Query Match	32.98;	Score 27.6;	DB 23;	Length.395;
Best Local Similarity	61.88;	Pred. No.75;		
Matches 42;	Conservative	0;	Mismatches 25;	Indels 0;
			Gaps	0

Query Match	32.9%;	Score 27.6;	DB 23;	Length 353;
Best Local Similarity	61.8%;	Pred. No. 75;		
Matches 42;	Conservative	0;	Mismatches 26;	Indels 0;
				Gaps 0

Query Match	32.98;	Score 27.6;	DB 32;	Length 202;
Best Local Similarity	61.88;	Pred. No. 80;		

	Matches	42;	Conservative	0;	Mismatches	26;	Indels	0;	Gaps
QY	17	AAACAACCA	CCACCTCGGTTAAATTGATGCAAAAATATGTCATCTACAGCATTTGAGGT	76					
Db	16	AAATAAACCA	CCACCTACAGTTAGATTGTGTTCTAAGATCTTCATCCAAATGTTATGCAGAT	75					
QY	77	AGCATCAT	84						
Db	76	GGTAGTAT	83						
RESULT	12								
LOCUS	AA398114/c								
DEFINITION	AA398114	430 bp	mRNA	EST	16-MAY-1997				
ACCSSION	AA398114								
NID	AA398114								
WORDS	92051223								
ORIGIN	AA398114.1	GI:2051223							
ORCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.								
	1 (bases 1 to 430)								
	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,								
	Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,								
	Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,								
	White, Y., Wylie, T., Waterson, R. and Wilson, R.								
TITLE	Washu-Merck EST Project 1997								
COMMENT	Unpublished (1997)								
	On Apr 14, 1993 this sequence version replaced q1:693347.								

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -4m13 fwd. ET from Amersham
 High quality sequence stop: 142.
 Location/Qualifiers
 1. .430

```

/organism="Homo sapiens"
/db_xref="GDB:5923466"
/ab_xref="taxon:9606"
/map="13"
/clone="IMAGE:726554"
/clone_id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc. and primed with a Not I - oligo(dT)
primer [5].
TGTACCAATCTGACAGTGGAGCGCGCCGCCCAATTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldc. "

```

Query Match	32.9%;	Score 27.6;	DB 33;	Length 430;
Best Local Similarity	60.8%;	Pred. No. 74;		
Matches 45; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0;

Oy	11	AATATGAAACCAACCACCTGGTTAAATTGATCGAAAAAATTTGCATCTCACAGCATTT	70
Dd	412	ATACCCATTAACACACCTTACAATTGTCTCTGAAGTGTCATCCAAAATGCTTAT	35
Oy	71	TCAAGTAGATCAT	84
Dd	352	GCAAGATGGTAGTAT	339

RESULT	13				
AA553441/c					
LOCUS	AA553441	493 bp	mrna	EST	12-AUG-1997
DEFINITION	nA77d07.s1 NC1 CGAL_Sch1 Homo sapiens cDNA clone IMAGE:1019533 3				
	similar to gb:M35663 INTERFERON-INDUCED, DOUBLE-STRANDED				
	RNA-ACTIVATED PROTEIN KINASE (HUMAN); mRNA sequence.				
ACCESSION	AA553441				
NID	92323980				
VERSION	AA553441.1	GI:2323980			
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 493)
AUTHORS	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997)
COMMENT	On Apr 14, 1993 this sequence version replaced g1:692959.

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-cGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbfp/image/image.html

```

FEATURES
source
1..493
Seq primer: -40m13 fwd. Err from Amersham
High quality sequence stop: 479.
Location/Qualifiers
1..493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="MMU16C3-C4 region; MMU16C3-C4 region"
/clone_image="1019533"
/clone_id="NCI_CGAP_Sch1"
/tissue_type="Schwannoma tumor"
/lab_host="SCLR (Kanamycin resistant)"
/note="Vector: Bluescript SK-; Site:1: EcoRI; Site:2:
XhoI; Cloned unidirectionally. Primer: Oligo dt 1wo
pooled bulk Schwannoma tumors. 5' adaptor sequence: 5',
GAATCGGACACGAG 3' adaptor sequence: 5',
CTCAGGATTTTATTTTATTTT 3' Average insert size: 1.2 kb."
152 a 97 c 72 g 172 t

```

Query Match	32.9%	Score 27.6;	DB 35;	Length 493;
Best Local Similarity	60.8%;	Pred. No. 73;		
Matches 45; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0

QY 7 AAGGTAATGAACACCACCATGGTTAAATTGTGCAAAAAATTGCATCTCACAGC 66
| | | | | | | | | | | | | | | |
Db 207 AAAGCAAAAAAAAAAGTACATTGGTGATTTAATGTAGATAATAATACATATGCCAAC 148
| | | | | | | | | | | | | | | |

QY 67 ATTTCAGSTAGA 80
| | | | | | | | | | | | | | | |

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:03:54 ; Search time 1962.17 Seconds

(without alignments)
108.594 Million cell updates/sec

Title: US-08-956-518a-98

Sequence: 1 TTTATCTGATGTCGCAATTC.....ACTGCATATGTAAGTACA 67

Scoring table: IDENTITY_NTC

Searched: 679419 seqs, 1590154680 residues

Abase :

```

GenEmb1:*
1: gb_da1:*
2: gb_da2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_hcg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_hcg1:*
35: gb_hcg2:*
36: gb_in1:*
37: gb_in2:*
38: em_da1:*
39: em_da2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	57	85.1	1836	11 AF037646 Homo sapi

2	49.8	74.3	64.1	11 AF029838	AF029838 Homo sapi
3	49.8	74.3	17.2	11 AF036903	AF036903 Homo sapi
4	47.4	70.7	689	11 AF029839	AF029839 Homo sapi
5	29.2	43.6	163403	37 AC007765	AC007765 Drosophila
6	28.8	43.0	167525	11 AC006531	AC006531 Homo sapi
7	26.8	40.0	7434	17 HPV9	X74464 Human papill
8	26.8	40.0	169891	35 AC007037	AC007037 Homo sapi
9	26.6	39.7	1084	12 M5MRNA	L36962 Mus musculu
10	25.8	38.5	100849	10 HUMEROF	L05367 Human Oligo
11	25.8	38.5	2638	12 AB015894	AB015894 Mus muscu
12	25.8	38.5	2638	12 AB015894	AB015894 Mus muscu
13	25.8	38.5	2638	12 AB015894	AB015894 Mus muscu
14	25.8	38.5	297898	42 AC004526	AC004526 Homo sapi
15	25.6	38.2	141313	11 AC004943	AC004943 Homo sapi
16	25.6	38.2	54705	34 AC006094	AC006094 Homo sapi
17	25.4	37.9	8393	1 U39732	U39732 Mycoplasma
18	25.4	37.9	9516	2 U39727	U39727 Mycoplasma
19	25.4	37.9	100625	10 HS1189K21	AL030997 Human DNA
20	25.4	37.9	89097	42 AC005518	AC005518 Homo sapi
21	25.4	37.9	28945	42 AF048729	AF048729 Homo sapi
22	25.2	37.6	74842	7 AB016877	AB016877 Arabidops
23	25.2	37.6	4030	7 POSTH4C11	M62755 Potato 4-co
24	25.2	37.6	140094	10 HS2013	AL035423 Human DNA
25	25.2	37.6	85530	34 HS469D22	AL031284 Homo sapi
26	25	37.3	32424	11 AC004037	AC004037 Homo sapi
27	24.8	37.0	183085	11 AC005815	AC005815, complet
28	24.6	36.7	1361	7 HSF21	D50797 Schizosacch
29	24.6	36.7	1361	7 HSF21	D50797 Schizosacch
30	24.6	36.7	42037	7 SPAC23D3	Y09057 H. annuus MR
31	24.6	36.7	81467	8 ATAC004683	Z64354 S. pombe chr
32	24.6	36.7	1522	36 TRBRCJ2G	AC004683 Arabidops
33	24.4	36.4	68098	7 AB020752	L42549 Trypanosoma
34	24.4	36.4	1975	12 M5MRNPK	AB020752 Arabidops
35	24.4	36.4	126983	12 M5MRNPK	L29769 House hnrNP
36	24.4	36.4	126983	12 M5MRNPK	L29769 House hnrNP
37	24.4	36.4	155313	8 AC006453	L11961 Mus cook11
38	24.2	36.1	6513	8 SIU55278	AC006453 Homo sapi
39	24.2	36.1	100000	9 AP000023	U55278 Solanum lyc
40	24.2	36.1	89086	9 AP000023	AP000023 Homo sapi
41	24.2	36.1	121290	34 CEH27C14	AP000023 Homo sapi
42	24.2	36.1	300172	35 AC005308	Z98852 Caenorhabd1
43	24.2	36.1	25282	36 CEC30F2	AC005308 Plasmodin
44	24.2	36.1	39353	36 CEF11A1	Z70681 Caenorhabd1
45	24.2	36.1	119996	42 AC006155	Z50857 Caenorhabd1

ALIGNMENTS

RESULT	1	AF037646	1836 bp	RNA	PRI	16-OCT-1998
LOCUS	AF037646	Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor precursor RNA, partial sequence.				
DEFINITION	AF037646					
ACCESSION	AF037646					
NID	93757808					
VERSION	AF037646.1	GI:3757808				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 1836)					
AUTHORS	Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, I., Jacobs, S., Meriwether, J., Choi, M. J., Kim, E. J., Walton, K., Bulting, R., Davis, A., Breese, C., Freedman, R. and Leonard, S.					
TITLE	Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7)					
JOURNAL	Genomics 52 (2), 173-185 (1998)					
MEDLINE	99000837					
REFERENCE	2 (bases 1 to 1836)					
AUTHORS	Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.					


```

TITLE      Direct Submission
JOURNAL    Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES   Location/Qualifiers
            source
              1..1836
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /db_xref="dbEST:M03952"
                /db_xref="dbEST:N73891"
                /chromosome="15"
                /clone="297134"
                /clone_1lb="Scores fetal liver spleen library, Research Genetics/IMAGE Consortium, LUNN"
                /note="unprocessed mRNA with intron"
          <1..1836
            /note="alpha-7 neuronal nicotinic acetylcholine receptor precursor; Intron included in 5' region"
BASE COUNT      383 a      503 c      469 g      481 t
ORIGIN
Query Match     85.1%; Score 57; DB 11; Length 1836;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TTTATTCTAGTTCACATTCGCTAATCCAGCATTTTGGAATAGCTGCANAACGTGGATAT 57
Db      160 TTTATTCTAGTTCACATTCGCTAATCCAGCATTTTGGAATAGCTGCANAACGTGGATAT 216
RESULT  2
LOCUS       AF029838                      641 bp      mRNA           PRI         16-DEC-1998
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
ACCESSION   AF029838
VERSION     93757793
SOURCE      AF029838.1 GI:37577793
ORGANISM    human.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 641)
AUTHORS     Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
             Moore,T., Jacobs,S., Melwether,J., Choi,M.J., Kim,E.J., Walton,K.,
             Bulling,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
TITLE       Genomic organization and partial duplication of the human alpha7
            neuronal nicotinic acetylcholine receptor gene
JURNAL      Genomics 52 (2), 173-185 (1998)
EDLINE      2 (bases 1 to 641)
REFERENCE   Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
AUTHORS     Breese,C., Davis,A., Hopkins,J. and Freedman,R.
TITLE       Direct Submission
JOURNAL     Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES   Location/Qualifiers
            source
              1..641
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="15"
                /map="15q14"
                /clone="CHRNA7-DR1"
                /note="RACE product A/B/C/D"
          <1..>641
            /note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"
BASE COUNT      151 a      192 c      163 g      135 t
ORIGIN
Query Match     74.3%; Score 49.8; DB 11; Length 641;
Best Local Similarity 96.2%; Pred No. 1.9e-07;

```

Matches	51: Conservative	0: Mismatches	2: Indels	0: Gaps	0:
QY	5	TTTCAGTCCAAATGCTAATCCAGCAATTTGTGATAGCTGCAAACTGGCATAT	57		
Db	482	TTTCAGTCCAAATGCTAATCCAGCAATTTGTGATAGCTGCAAACTGGCATAT	534		
RESULT	3				
AF036903		1712 bp	mRNA	PRI	16-OCT-1998
LOCUS		Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor			
DEFINITION		mRNA, alternatively spliced, partial sequence.			
ACCESSION		AF036903			
NID		93737807			
VERSION		AF036903.1	GI:37577807		
KEYWORDS		human.			
SOURCE		Homo sapiens			
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
REFERENCE		Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS		1 (bases 1 to 1712)			
TITLE		Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,			
JOURNAL		Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,			
MEDLINE		Building,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.			
REFERENCE		Genomic organization and partial duplication of the human alpha7			
AUTHORS		neural nicotinic acetylcholine receptor gene (CHRNA7)			
TITLE		Genomics 52 (2), 173-185 (1998)			
JOURNAL		2 (bases 1 to 1712)			
MEDLINE		Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,			
REFERENCE		Breese,C., Davis,A., Hopkins,J. and Freedman,R.			
AUTHORS		Direct Submission			
TITLE		Submitted (03-DEC-1997) Psychiatry, University of Colorado Health			
JOURNAL		Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA			
FEATURES		Location/Qualifiers			
SOURCE		1..1712			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/db_xref="dbEST:W52861"			
		/db_xref="dbEST:AA037389"			
		/chromosome="15"			
		/clone="321242"			
		/clone_1lb="Scars parathyroid tumor NbHRA library,			
		Research Genetics/IMAGE Consortium, LNLV"			
		<1..1712			
		/note="alpha-7 neuronal nicotinic acetylcholine receptor;			
		alternatively spliced"			
misc_feature					
BASE COUNT		356 a 489 c 457 g 410 t			
ORIGIN					
Query Match		74.3%; Score 49.8; DB 11; Length 1712;			
Best Local Similarity		96.2%; Pred. No. 1.8e-07;			
Matches	51: Conservative	0: Mismatches	2: Indels	0: Gaps	0:
QY	5	TTTCAGTCCAAATGCTAATCCAGCAATTTGTGATAGCTGCAAACTGGCATAT	57		
Db	42	TTTCAGTCCAAATGCTAATCCAGCAATTTGTGATAGCTGCAAACTGGCATAT	94		
RESULT	4				
AF029839		689 bp	mRNA	PRI	16-DEC-1998
LOCUS		Homo sapiens alpha 7 neuronal nicotinic acetylcholine receptor mRNA sequence.			
DEFINITION		AF029839			
ACCESSION		93757794			
NID		AF029839.1	GI:3757794		
VERSION		human.			
KEYWORDS		Homo sapiens			
SOURCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
ORGANISM		Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE		1 (bases 1 to 689)			

TITLE	JOURNAL
<p>Direct Submission Submitted (09-JUN-1999) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US On Jun 9, 1999, this sequence version replaced gi:3293203.gi:3097820.</p>	
COMMENT	<p>Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bogd@fruitfly.berkeley.edu. PI library locations: 23-78, 10-42. Location/Qualifiers</p>
FEATURES	<p>source</p> <p>1..163403 /organism="Drosophila melanogaster" /strain="y; cn br sp" /db_xref="taxon:7227" /chromosome="2L" /map="23C1-23C5" /clone_lib="pi library, partial Sau3A in pAG10secBIR" /clone="pis DS02190 (DB2) and DS00906 (DS9)" /note=""</p>
BASE COUNT	<p>45206 a 35896 c 36551 g 45750 t</p>
ORIGIN	
Query Match	<p>43.6%; Score 29.2; DB 37; Length 163403; Best Local Similarity 65.2% ; Pred. No. 1.4;</p>
Matches	<p>43; Conservative 0; Mismatches 23; Indels 0; Gaps 0</p>
OY	<p>2 TTATCTGATTCCACTTGTCTATAACCATTTGGTAGTCGAACAACGCATATGTAA 61 Db 108123 TTAATATCTTAATTTATTACTCGCGCAITTGTAATTAGTCCCAAATGAGTAATATA 108182</p>
OY	<p>62 GTACA 67 Db 108183 ATCAC 108188</p>
RESULT	<p>6 AC006531 DNA PRI 06-FEB-1999</p>
LOCUS	<p>AC005531 167525 bp complete sequence.</p>
DEFINITION	<p>Homo sapiens chromosome 16 clone 113K5,</p>
ACCESSION	<p>AC006531 g4235137</p>
NID	<p>AC006531.1 GI:4235137</p>
VERSION	<p>HTG.</p>
KEYWORDS	<p>human.</p>
SOURCE	<p>Homo sapiens</p>
ORGANISM	<p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.</p>
REFERENCE	<p>AUTHORS</p> <p>(bases 1 to 167525) Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatam,O., Campbell,C., Fawcett,J., Malbtie,M., Bussod,M., Sutherland,R., Han,C. and Deaven,L.</p>
TITLE	<p>Sequencing of Human Chromosome 16p13.3 Unpublished</p>
JOURNAL	<p>2 (bases 1 to 167525)</p>
REFERENCE	<p>Ricke,D.O.</p>
AUTHORS	<p>LARGE Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System</p>
TITLE	<p>Unpublished</p>
JOURNAL	<p>3 (bases 1 to 167525)</p>
REFERENCE	<p>Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatam,O.,</p>
AUTHORS	<p>Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatam,O.,</p>

Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.

TITLE Direct Submission
JOURNAL Submitted (06-FEB-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA

FEATURES Location/Qualifiers
1. 167525

repeat_region /organism="Homo sapiens"
misc_feature /db_xref="taxon:9606"
/clone="113K5"
/chromosome="16"
2396..2466
/rpt_family="MIR"
complement(join(3760..10746,14646..14885,44104..44260,
68059..68223,75207..75450,79902..80033,86444..86599,
86723..86902,95554..95760))
/note="NMDA receptor modulatory subunit"
12465..12543
/rpt_family="Alu"
complement(13201..13294)
/rpt_family="L1"
repeat_region complement(13836..14146)
/rpt_family="Alu"
complement(14600..14851)
/note="GRAIL 2 excellent exon, frame 2"
complement(16599..16853)
/rpt_family="L1"
repeat_region complement(18201..18458)
/rpt_family="Alu"
23577..23860
/rpt_family="Alu"
complement(24365..25041)
/rpt_family="MER"
26799..26960
/rpt_family="MER"
repeat_region complement(27162..27453)
/rpt_family="Alu"
28101..28204
/rpt_family="MER2"
29619..29862
/rpt_family="L1"
29943..30269
/rpt_family="Alu"
30282..31394
/rpt_family="L1"
32685..32828
/rpt_family="MIR"
33206..33361
/rpt_family="MER4"
complement(34052..34381)
/rpt_family="Alu"
complement(34520..34772)
/rpt_family="Alu"
35856..36026
/note="GRAIL 2 excellent exon, frame 1"
complement(38026..38318)
/rpt_family="Alu"
38718..38821
/rpt_family="MER5"
complement(40648..40860)
/note="GRAIL 2 excellent exon, frame 1"
complement(40863..41382)
/rpt_family="MER41"
complement(43625..43725)
/rpt_family="MER20"
complement(44039..44226)
/note="GRAIL 2 excellent exon, frame 2"
45220..45336
/rpt_family="MLT"
45337..45419
/rpt_family="MER"
complement(45749..46003)

misc_feature /rpt_family="Alu"
46499..46629
/note="GRAIL 2 excellent exon, frame 0"
47808..48246
/rpt_family="L1"
48509..48629
/rpt_family="MER3"
48919..49218
/rpt_family="Alu"
complement(49485..49635)
/rpt_family="MIR"
51405..51488
/rpt_family="MIR"
52326..52469
/note="GRAIL 2 excellent exon, frame 1"
52626..52963
/rpt_family="THE1"
complement(53966..54143)
/rpt_family="MIR"
complement(54381..54442)
/rpt_family="MIR"
54933..55231
/rpt_family="Alu"
55579..55875
/rpt_family="Alu"
56250..56595
/rpt_family="MLT1"
57272..57582
/rpt_family="Alu"
58153..58480
/rpt_family="Alu"
60534..60842
/rpt_family="Alu"
61134..61527
/rpt_family="L1"
61633..61910
/rpt_family="L1"
61958..62058
/rpt_family="L1"
62540..62679
/rpt_family="MER5"
64727..64830
/rpt_family="L1"
65034..65336
/rpt_family="Alu"
65464..65742
/rpt_family="Alu"
complement(68027..68187)
/note="GRAIL 2 excellent exon, frame 1"
complement(69367..69526)
/rpt_family="MIR"
complement(75186..75415)
/note="GRAIL 2 excellent exon, frame 2"
complement(75606..75636)
/note="GRAIL 2 excellent exon, frame 2"
complement(77284..77625)
/rpt_family="Alu"
complement(77735..77896)
/rpt_family="Alu"
complement(77916..78047)
/rpt_family="Alu"
complement(79531..79812)
/rpt_family="Alu"
complement(79868..79993)
/note="GRAIL 2 excellent exon, frame 2"
complement(80168..80463)
/rpt_family="Alu"
84026..84320
/rpt_family="L1"
84674..84820
/rpt_family="L1"
85690..85917
/rpt_family="Alu"


```

gene      VMSLGNVDL"
4129. .5730
/ gene="L2"
4129. .5730
/ gene="L2"
/ codon_start=1
/ product="late protein"
/ protein_id="CA52487.1"
/ db_xref="PID:9397075"
/ db_xref="GI:397075"
/ translation="MYRARRTRASTYTDYIRGCAAGTCPEPVINKVENTHTIADRILO
VSAAGVFRGGLGISTGRTGATGYVPLEGPGVAGVPTIIRPGVITPEIIGPDLI
PLOTVAIPDIPAPASIVTGTVDLPGIEISIAEHPVNAVAVDVPVEGRRGS
SAILEVADPAPMRETRARTQXHNPAFOIISSTPMGSSIAADHIIVEGSGCOLY
GRESYTASSENIELOEPPRSYSFEIDESTPTSTPORAOSLSLRATYKRLT
EOVAAYDPLFLSRSLNOFODNPAFEDVEVQIIEBRLSTVEEPPDRQFLDVLRLR
PLVETPGVYSRIRGRATRTISGAGVGAQVHFYRDLSTINTEEPLEMOLEHS
GSTIVQGVESIVDNIPEDEPGLVGRQETSVEDNEDSLLDEGVDEFSGQ
LVGTRSTNTLTVPRFETPDTSTFYIDIQYTVSYSPESRQTDIIIFPHDPTIVVI
HNDISGYLHPSLQRRKKRKYL"
/ gene="L1"
5745. .7268
/ gene="L1"
/ codon_start=1
/ product="late protein"
/ protein_id="CA52488.1"
/ db_xref="PID:9397076"
/ db_xref="GI:397076"
/ db_xref="SWISS-PROT:P36746"
/ translation="MSLWLPASGKYILPPATPVAVOSTDEYVERTNIEYHAISDRLL
TVGHHYVYRSGDGRLEVPKSYNOYAFRISLPDRFALAMSYNPNKQELTYA
CRGIEIGGOLGIGTSGPLFNKYRDSNNOGTMDRQNTSDPKQKFTIG
CIPCGEHMDKAKYCCENANNOLGICPTELRNTVIEDGDFDIFGFINNKEISFNK
SDVSDIYDERCKYDFLTMANDYGDACCFEARECOYAHYVYRSGVDAVPDGA
VNOHNFPLPAKSDQOORTIYNSSTYPSGLVTSAPOLFNRPEFMOIRAGHNGIL
WGNQIFVADNTRNTFTISTEAOQEEYNNMNTREYLRHVEFOISLIIOLCY
PLVLEVISQINAMNSGILEDMOLGVPVPEPAVDIYRIDSATKCPDAVEPEREKD
PFAYSPKVKVLTRELSDLOXPGRFLQACGLQTRKRIKTSVATSKAKARRR"
BASE COUNT      2363 a      1393 c      1654 g      2024 t
ORIGIN

Query Match      40.0%; Score 26.8; DB 17; Length 7434;
Best Local Similarity 64.5%; Pred. No. 12;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

1 TTTATTCAGTTCATTCATTCAGCATTGTGATAGCTGCAACTGCATATGTA 60
|| || || || || || || || || || || || || || || || || || ||
Db 5005 TTAGTTCATTCAGTTCATTCAGCATTGTGATAGCTGCAACTGCATATGTA 5064

OY 61 AG 62
||
Db 5065 AG 5066

RESULT 8
AC007037 169891 bp DNA HTG 07-APR-1999
LOCUS Homo sapiens clone NH0119H15, WORKING DRAFT SEQUENCE, 6 unordered
pieces.
AC007037
AC007037
VERSION 94572700
AC007037.2 GI:4572700
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169891)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169891)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted. (06-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 7, 1999 this sequence version replaced gi:4371308.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2135: contig of 2135 bp in length
* 2136 2153: gap of unknown length
* 2154 7671: contig of 3518 bp in length
* 7672 7689: gap of unknown length
* 7690 22866: contig of 15176 bp in length
* 22867 22884: gap of unknown length
* 22885 56499: contig of 33616 bp in length
* 56500 56517: gap of unknown length
* 56518 102994: contig of 46477 bp in length
* 102995 103012: gap of unknown length
* 103013 169891: contig of 66879 bp in length.
Location/Qualifiers
1. 169891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0119H15"
BASE COUNT      50240 a      35106 c      35944 g      48511 t      90 others
ORIGIN

Query Match      40.0%; Score 26.8; DB 35; Length 169891;
Best Local Similarity 64.5%; Pred. No. 9.1;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 5 TTTAGTTCATTCATTCAGCATTGTGATAGCTGCAACTGCATATGTA 64
|| || || || || || || || || || || || || || || || || || ||
Db 42387 TTATGAGCACATGACCTGACCTGATTCCTGTGAGCTGATGTGAGATGTA 42328

OY 65 AC 66
||
Db 42327 AC 42326

RESULT 9
MUSMENA 1084 bp mRNA ROD 03-DEC-1998
LOCUS Mus musculus metaxin mRNA, complete cds.
AC007037
AC007037
VERSION 9807669
AC007037.1 GI:807669
KEYWORDS glucocerebrosidase; metaxin; thrombospondin 3.
SOURCE Mus musculus (strain C57BL/6) brain cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bornstein,P., McKinney,C.E., Lamarca,M.E., Winfield,S., Shingu,T.,
Devareyalu,S., Vos,H.L. and Glans,E.I.
Metaxin, a gene contiguous to both thrombospondin 3 and
glucocerebrosidase, is required for embryonic development in the
mouse: Implications for Gaucher disease
Proc. Natl. Acad. Sci. U.S.A. 92 (10), 4547-4551 (1995)
95273400
Location/Qualifiers
1. 1084
/organism="Mus musculus"
/strain="C57BL/6"
JOURNAL MEDLINE
FEATURES source

```



```

TITLE      Brownstein,B.H. and Collins,F.
JOURNAL    A de novo Alu insertion results in neurofibromatosis type 1
MEDLINE    Nature 353 (6347), 864-866 (1991)
REFERENCE  6 (bases 1 to 100849)
AUTHORS    Weiss,R.B., Dunn,D., Disera,L., Wheatley,W., Kimball,A., Rote,C.,
            Cherry,J., Duval,B., Lee,R., Ferguson,M.W.J. and Gesteland,R.F.,
            The Human Neurofibromatosis Type 1 locus: genomic sequence of the
            3' region
TITLE      Unpublished (1992)
JOURNAL     7 (bases 1 to 100849)
REFERENCE   Xu,G., O'Connell,P., Stevens,J. and White,R.
AUTHORS     Characterization of human adenylylate kinase 3 (AK3) cDNA and mapping
            of the AK3 pseudogene to an intron of the NF1 gene
JOURNAL     Genomics 13 (3), 537-542 (1992)
MEDLINE    92347846
COMMENT     Submitting author Robert Weiss may be contacted at the following
            address and telephone number:
            address: Room 7160, Eccles Institute of Human Genetics, University of Utah, Salt
            Lake City, Utah 84112
            telephone: (801) 585-3436.
            e-mail:
            1. 100849
            location/Qualifiers
FEATURES
source      1. 100849
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="17q11.2"
            <1. 36919
            /number=1
            /complement(5761..8342)
            /gene="OMG"
            /complement(5761..7447)
            /gene="OMG"
            /note="G00-127-563"
            /citation=[4]
            /number=2
            /complement(7448..8265)
            /gene="OMG"
            /note="G00-127-563"
            /number=1
            /complement(8266..8342)
            /gene="OMG"
            /note="G00-127-563"
            /citation=[4]
            /number=1
            /complement(1223..1224)
            /rpt_family="GT"
            /complement(14658..16737)
            /gene="EVI2B"
            /note="G00-125-901"
            /citation=[3]
            /number=2
            /complement(14658..25133)
            /gene="EVI2B"
            /complement(16738..25076)
            /gene="EVI2B"
            /note="G00-125-901"
            /number=1
            /complement(24830..24856)
            /rpt_family="GT"
            /complement(25077..25133)
            /gene="EVI2B"
            /note="G00-125-901"
            /citation=[3]
            /number=1
            /complement(28465..28503)
            /rpt_family="AC"
            /complement(28542..30121)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=2
            /complement(28542..32788)
            /gene="EVI2A"
            /number=33

Intron      complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

exon        complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

mutation    complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

Intron      complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

exon        complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

mutation    complement(30122..32588)
            /gene="EVI2A"
            /note="G00-125-191"
            /number=1
            /complement(32589..32788)
            /gene="EVI2A"
            /note="G00-125-191"
            /citation=[2]
            /number=1
            /complement(33193..33232)
            /rpt_family="AC"
            /complement(33232..36920)
            /gene="NF1"
            /note="36920 = nt. 5032 from full-length neurofibromin
            cDNA (HUMNF1B,M89914); G00-120-231"
            /number=28
            /complement(36920..86010)
            /gene="NF1"
            /complement(36920..38596)
            /gene="NF1"
            /note="G00-120-231"
            /number=28
            /complement(38597..38937)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(38938..41643)
            /gene="NF1"
            /note="G00-120-231"
            /number=29
            /complement(41644..41846)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(41847..46170)
            /gene="NF1"
            /note="G00-120-231"
            /number=30
            /complement(46171..46364)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(46365..47772)
            /gene="NF1"
            /note="G00-120-231"
            /number=31
            /complement(47773..47913)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(47914..48075)
            /gene="NF1"
            /note="G00-120-231"
            /number=32
            /complement(48076..48355)
            /gene="NF1"
            /note="G00-120-231"
            /number=33

```



```

JOURNAL REFERENCE AUTHORS
    2 (bases 1 to 297898)
    vonNiederhausen,A.
TITLE JOURNAL
    Submitted (04-APR-1998) Human Genetics, University of Utah, 20 S.
REFERENCE AUTHORS
    3 (bases 1 to 297898)
    Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
    Hamil,C., Holmes,C., Korzeniewski,G., Mahmoud,M., Rose,R.,
    Stokes,R., Wang,C., Yu,P., Zhou,L., Gltin,Y., Nelson,J. and
    vonNiederhausen,A.
TITLE JOURNAL
    Direct Submission
    Submitted (17-OCT-1998) Human Genetics, University of Utah, 20 S.
REFERENCE AUTHORS
    4 (bases 1 to 297898)
    Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
    Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
    Yu,P., Zhou,L., Gltin,Y., Nelson,J. and vonNiederhausen,A.
TITLE JOURNAL
    Submitted (17-OCT-1998) Human Genetics, University of Utah, 20 S.
REFERENCE AUTHORS
    5 (bases 1 to 297898)
    Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
    Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
    Yu,P., Zhou,L., Gltin,Y., Nelson,J. and vonNiederhausen,A.
TITLE JOURNAL
    Direct Submission
    Submitted (12-OCT-1998) Human Genetics, University of Utah, 20 S.
REFERENCE AUTHORS
    5 (bases 1 to 297898)
    Weiss,R.B., Dunn,D.M., Aoyagi,A., Banks,L., Disera,L., Duval,B.,
    Hamil,C., Holmes,C., Mahmoud,M., Rose,R., Stokes,R., Stump,D.,
    Yu,P., Zhou,L., Gltin,Y., Nelson,J. and vonNiederhausen,A.
COMMENT JOURNAL FEATURES
    Submitted (25-FEB-1999) Human Genetics, University of Utah, 20 S.
    On Oct 22, 1998 this sequence version replaced g1:3766112.
    Location/Qualifiers
        1. 297898
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="23013,14502"
            /chromosome="21"
            /clone_1lb="CITB-978SR-B"
BASE COUNT      82446 a 60770 c 62692 g 91988 t      2 others
ORIGIN
Query Match          38.5%   Score 25.8; DB 42; Length 297898;
Best Local Similarity 63.9%;   Pred. No. 19;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0
1 TTTATTGATTCATCCATTCCTAATCATCAACAGATTGGTGATAGCTGCAAATCGCATATGTGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
188378 TCTTATTAGTTAGTGTATCTTAATCTGCGAATTAAAGGATTAACAGACGACTGTCTATTTC 188319
OY      61 A 61
Db 188318 A 188318
RESULT 15 AC004943/C
LOCUS AC004943 141313 bp DNA PRI 24-NOV-1998
DEFINITION Homo sapiens PAC clone DIO991G20, complete sequence.
ACCESSION AC004943
NID G3924671
VERSION AC004943.1 GI:3924671
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 141313)
Kozlowicz,A., McQuerry,Y. and Hottic,M.
The sequence of Homo sapiens PAC clone DIO991G20
Unpublished (1998)
2 (bases 1 to 141313)

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

Waterston, R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
3 (bases 1 to 141313)
Waterston, R.
Direct Submission
Submitted (24-NOV-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
ON Nov 24, 1998 this sequence version replaced g1:33264783.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:saplens@watson.wustl.edu

COMMENT

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pleier de Jong and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pleier de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of D0991G20;
actual end is at 141313 of D0991G20

FEATURES
Source
The location of this clone is unknown.
Location/Qualifiers
1..141313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="unknown"
/clone_id="RPCI-5"
/clone="D0991G20"
30..152
/rpt_family="retroviral"
405..634
/rpt_family="MIR"
686..865
/rpt_family="MER2_type"
866..1014
/rpt_family="MIR"
1047..1185
/rpt_family="MIR"
1947..1969
/rpt_family="MIR"
3058..3184
/rpt_family="AT-rich"
3447..4731
/rpt_family="Alu"
/note="cytokeratin pseudogene"
4869..5017
/rpt_family="L2"
5039..5247
repeat_region
repeat_region
misc_feature
repeat_region
repeat_region
repeat_region

```



```

repeat_region      5283. .5467      /rpt_family="L2"
repeat_region      5512. .11701     /rpt_family="L1"
repeat_region      11826. .12139     /rpt_family="Alu"
repeat_region      12557. .12585     /rpt_family="Alu"
repeat_region      12586. .12704     /rpt_family="BC200"
repeat_region      13614. .13642     /rpt_family="Alu"
repeat_region      13655. .13952     /rpt_family="AT_rich"
repeat_region      14871. .14956     /rpt_family="Alu"
repeat_region      14973. .15007     /rpt_family="L2"
repeat_region      15010. .15149     /rpt_family="AT_rich"
repeat_region      15159. .15429     /rpt_family="Alu"
repeat_region      15850. .15984     /rpt_family="L2"
repeat_region      16009. .16299     /rpt_family="MIR"
repeat_region      16840. .16947     /rpt_family="Alu"
repeat_region      17292. .17343     /rpt_family="purine-rich"
repeat_region      17684. .18442     /rpt_family="L2"
repeat_region      18523. .18544     /rpt_family="L1"
repeat_region      19220. .19245     /rpt_family="AT_rich"
repeat_region      20355. .20789     /rpt_family="AT_rich"
repeat_region      20862. .20991     /rpt_family="L1"
repeat_region      21053. .21172     /rpt_family="MER94"
repeat_region      21409. .21712     /rpt_family="MER1_type"
repeat_region      22318. .22613     /rpt_family="Alu"
repeat_region      23077. .23206     /rpt_family="Alu"
repeat_region      24150. .24824     /rpt_family="Alu"
repeat_region      25528. .25966     /rpt_family="L1"
repeat_region      26005. .26113     /rpt_family="Retroviral"
repeat_region      26125. .26800     /rpt_family="Alu"
repeat_region      26801. .27101     /rpt_family="Retroviral"
repeat_region      27102. .27431     /rpt_family="Alu"
repeat_region      27433. .27560     /rpt_family="Retroviral"
repeat_region      27643. .27725     /rpt_family="L1"
repeat_region      27726. .28016     /rpt_family="Alu"
repeat_region      28017. .28510     /rpt_family="Alu"
repeat_region      28542. .28617     /rpt_family="L1"
repeat_region      28676. .29224     /rpt_family="AT_rich"

```

```

repeat_region      29558. .29851     /rpt_family="MALR"
repeat_region      30066. .30242     /rpt_family="Alu"
repeat_region      30469. .30764     /rpt_family="MALR"
repeat_region      30856. .31238     /rpt_family="Alu"
repeat_region      31239. .31306     /rpt_family="L1"
repeat_region      31308. .31970     /rpt_family="(CAT)n"
repeat_region      31971. .32271     /rpt_family="L1"
repeat_region      32272. .32324     /rpt_family="Alu"
repeat_region      32344. .32683     /rpt_family="L1"
repeat_region      32659. .32721     /rpt_family="L1"
repeat_region      32741. .32865     /rpt_family="L1"
repeat_region      32876. .34100     /rpt_family="Alu"
repeat_region      34101. .34374     /rpt_family="L1"
repeat_region      34375. .35747     /rpt_family="Alu"
repeat_region      35735. .36153     /rpt_family="L1"
repeat_region      36154. .36701     /rpt_family="MALR"
repeat_region      36852. .37128     /rpt_family="MALR"
repeat_region      37644. .38038     /rpt_family="L1"
repeat_region      40126. .40182     /rpt_family="MALR"
repeat_region      /rpt_family="Alu"

```

```

Query Match      38.2%; Score 25.6; DB 11; Length 141313;
Best Local Similarity 66.1%; Pred. No. 24;
Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY      1 TTATTCCTAGTCCATTCCTAATCCAGCATTTGTGATAGCTGCATACTGGATA 56
Db 90935 TTCTCTAGTCTTTTCTTCTAATACCTACTTTTACATAGCTGTAACATGGATA 90880

```

Search completed: September 17, 1999, 22:04:19
Job time: 16442 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:20 ; Search time 2825.05 Seconds

(without alignments)
46.781 Million cell updates/sec

Title: US-08-956-518a-98

Perfect score: 67
Sequence: 1 TTTTCTGATGTCGATTC.....ACTGCGATGCTAGTACA 67

Scoring table: IDENTITY_NDC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:
1: em_est1:
2: em_est2:
3: em_est3:
4: em_est4:
5: em_est5:
6: em_est6:
7: em_est7:
8: em_est8:
9: em_est9:
10: em_est10:
11: em_est11:
12: em_est12:
13: em_est13:
14: em_est14:
15: em_est15:
16: em_est16:
17: em_est17:
18: em_est18:
19: em_est19:
20: em_est20:
21: em_est21:
22: em_est22:
23: em_est23:
24: em_est24:
25: em_est25:
26: em_est26:
27: em_est27:
28: em_est28:
29: em_est29:
30: em_est30:
31: em_est31:
32: em_est32:
33: em_est33:
34: em_est34:
35: em_est35:
36: em_est36:
37: em_est37:
38: em_est38:
39: em_est39:
40: em_est40:
41: em_est41:
42: em_est42:
43: em_est43:
44: em_est44:
45: em_est45:
46: em_est46:
47: em_est47:
48: em_est48:
49: em_est49:
50: em_est50:
51: em_est51:
52: em_est52:
53: em_est53:

54: em_est22:
55: em_est23:
56: em_est24:
57: em_est25:
58: em_est26:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48.6	72.5	487	39	AA861176	AA861176 AK36009.s
2	48.2	71.9	560	26	W52861	W52861 zc03e02.r1
3	44.4	65.7	409	25	W03952	W03952 za62c08.r1
4	43.2	64.5	274	34	AA460018	AA460018 z49h09.s
5	43.2	64.5	436	39	AA844642	AA844642 a357a12.s
6	43.2	64.5	396	45	AI367541	AI367541 qy93c12.x
7	43.2	64.5	501	49	AI655990	AI655990 tt42c03.x
8	27.7	40.3	354	26	W85937	W85937 zh60e03.r1
9	26.6	39.7	560	26	W76900	W76900 me58b02.r1
10	26.6	39.7	416	26	W85213	W85213 m51b09.r1
11	26.6	39.7	517	27	W89656	W89656 m81f04.r1
12	26.6	39.7	590	30	AA222672	AA222672 mv99e05.r
13	26.6	39.7	530	33	AA388232	AA388232 vb59c01.r
14	26.6	39.7	545	35	AA590415	AA590415 vml6d07.r
15	26.6	39.7	479	37	AA719305	AA719305 zb35c02.s
16	26.6	39.7	499	38	AA771094	AA771094 vt16e07.r
17	26.6	39.7	542	38	AA792253	AA792253 vn94e09.r
18	26.6	39.7	618	40	C89142	C89142 C89142 Mous
19	26.6	39.7	1048	44	AI255545	AI255545 u155h01.y
20	25.6	38.2	291	20	T05021	T05021 EST02909.Fe
21	25.2	37.6	445	47	AI488647	AI488647 EST246986
22	25.2	37.6	451	47	AI488637	AI488637 EST247176
23	24.6	36.7	467	42	AI099847	AI099847 34000.Lam
24	24.6	36.7	347	43	C99808	C99808 C99808 YAC
25	24.6	36.7	297	48	AI555647	AI555647 UI-R-C2P-
26	24.4	36.4	633	25	W10059	W10059 me67d04.r1
27	24.4	36.4	610	27	AA027731	AA027731 m14e01.r
28	24.4	36.4	669	27	AA027739	AA027739 m14g01.r
29	24.4	36.4	611	27	AA033365	AA033365 m143d06.r
30	24.4	36.4	582	27	AA048492	AA048492 m30c08.r
31	24.4	36.4	627	28	AA061156	AA061156 m31c08.r
32	24.4	36.4	466	28	AA103275	AA103275 mo25h04.r
33	24.4	36.4	420	28	AA107857	AA107857 mo49h08.r
34	24.4	36.4	322	28	D76888	D76888 MDS91F01.mo
35	24.4	36.4	654	29	AA170084	AA170084 ms32g06.r
36	24.4	36.4	503	29	AA183839	AA183839 mo95a04.r
37	24.4	36.4	576	29	AA184588	AA184588 m51c02.r
38	24.4	36.4	598	30	AA197755	AA197755 mv03a06.r
39	24.4	36.4	490	34	AA517605	AA517605 v933d04.r
40	24.4	36.4	616	35	AA544863	AA544863 vk38b07.r
41	24.4	36.4	500	35	AA561216	AA561216 v134e02.r
42	24.4	36.4	554	37	AA711653	AA711653 vu27b09.r
43	24.4	36.4	828	38	AA755792	AA755792 vv35d03.r
44	24.4	36.4	555	38	AA794314	AA794314 vu77e08.r
45	24.4	36.4	880	50	AU066965	AU066965 AU066965

ALIGNMENTS

RESULT 1
AA861176/c
LOCUS
DEFINITION AK36009.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1408049
ACCESSION AA861176
NID 92953316
VERSION AA861176.1 GI:2953316

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 487)
JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 9, 1998 this sequence version replaced gi:937098.

FEATURES
source
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html

Insert Length: 758 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 458.
 Location/Qualifiers
 1..487
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1408049"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech.
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTTCACATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cots, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 95 c 102 g 159 t

GIN

Query Match 72.5%; Score 48.6; DB 39; Length 487;
 Best Local Similarity 92.7%; Pred. No. 3.1e-07;
 Matches 51; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 5 TTTAGTTCATTGCTATCCAGCATTTGTGATGATGCAACTGGCATATG 59
 11 ||||||||||||||||||||||||||||||||||||||||
 462 TTTAGTTCATTGCTATCCAGCATTTGTGATGATGCAACTGGCATATG 408

RESULT 2
LOCUS W52861 560 bp mRNA EST 10-OCT-1996
DEFINITION zc03e02.r1 Soares-parathyroid tumor_NBHPA Homo sapiens cDNA clone
 IMAGE:321242 5', similar to gb:X70297 NEURONAL ACETYLCHOLINE
 RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION W52861
NID g1350351
VERSION W52861.1 GI:1350351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 560)

AUTHORS
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
TITLE The Wash-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Jan 24, 1995 this sequence version replaced gi:634363.

FEATURES
source
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1753 Std Error: 0.00
 Seq primer: mob.RS6A-ET
 High quality sequence stop: 408.
 Location/Qualifiers
 1..560
 /organism="Homo sapiens"
 /db_xref="cds:1258800"
 /db_xref="taxon:9606"
 /clone="IMAGE:321242"
 /clone_lib="Soares-parathyroid tumor_NBHPA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: parathyroid gland; Vector: pT73D
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 [5',
 TGTTCACATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was size selected, ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library went through one round of
 normalization to a Cots = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
 adenomas was kindly provided by Dr. Stephen Marx, National
 Institute of Diabetes and Digestive and Kidney Diseases,
 NIH."

BASE COUNT 128 a 137 c 137 g 152 t 6 others

ORIGIN

Query Match 71.9%; Score 48.2; DB 26; Length 560;
 Best Local Similarity 94.3%; Pred. No. 4.3e-07;
 Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 5 TTTAGTTCATTGCTATCCAGCATTTGTGATGATGCAACTGGCATATG 57
 11 ||||||||||||||||||||||||||||||||||||||||
 34 TTTAGTTCATTGCTATCCAGCATTTGTGATGATGCAACTGGCATATG 86

RESULT 3
LOCUS W03952 409 bp mRNA EST 19-APR-1996
DEFINITION za62e08.r1 Soares fetal liver spleen INFES Homo sapiens cDNA clone
 IMAGE:297114 5', similar to gb:X70297 NEURONAL ACETYLCHOLINE
 RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION W03952
NID g1275817
VERSION W03952.1 GI:1275817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 409)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE
JOURNAL
COMMENT
The Mashu-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET

FEATURES

SOURCE

High quality sequence stop: 263.
Location/Qualifiers

1. 409
/organism="Homo sapiens"
/db_xref="GDB:1242055"
/db_xref="taxon:9606"
/clone="IMAGE:297134"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AAGCGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

95 a 80 c 75 g 145 t 14 others

Query Match 65.7%; Score 44; DB 25; Length 409;
Best Local Similarity 94.8%; Pred. No. 1.2e-05;

Matches 55; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1 TTTATCTAGTTCATGCTTAATCCA-GCATTGTGATGCTGCAACGCGATAT 57
|||||
Db 149 TTTATCTAGTTCATGCTTAATCCA-GCATTGTGATGCTGCAACGCGATAT 206

AA460018 274 bp mRNA EST 09-JUN-1997
LOCUS AA460018/c
DEFINITION zyx9b09.s1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:195617
ACCESSION AA460018
NID 92184902
VERSION AA460018.1 GI:2184902
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 274)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marr, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

AUTHORS

TITLE
JOURNAL
COMMENT
Washu-Merck EST Project 1997
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced g1:1393651.

Contact: Wilson RK

FEATURES

SOURCE

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amersham.

Location/Qualifiers

1. 274
/organism="Homo sapiens"
/db_xref="GDB:6038961"
/db_xref="taxon:9606"
/clone="IMAGE:795617"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTCCATCTGATGAGGAGGCGCCGCGCCCATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

78 a 53 c 67 g 76 t

Query Match 64.5%; Score 43.2; DB 34; Length 274;
Best Local Similarity 93.8%; Pred. No. 2.2e-05;

Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 10 GTTCCATGCTTAATCCAGCATTTGTGATGCTGCAACGCGATAT 57
|||||
Db 256 GTTCCATGCTTAATCCAGCATTTGTGATGCTGCAACGCGATAT 209

AA844642 436 bp mRNA EST 31-DEC-1998
LOCUS AA844642/c
DEFINITION a357a12.s1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1394398
ACCESSION AA844642
NID 92931093
VERSION AA844642.1 GI:2931093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 436)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/cgi/cgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index.

AUTHORS

TITLE
JOURNAL
COMMENT
Unpublished (1997)
On Jan 17, 1998 this sequence version replaced g1:1900948.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 497 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 415.
Location/Qualifiers
1. 436

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1: 1"
/clone="IMAGE:1394398"
/clone.lib="Soares_testis_NH"
/sex="male"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer 15',
TCTTACCAATTCGATCAGACATTGTGATGCTGCAACTGGCATAT 3').
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 115 a 96 c 103 g 122 t
ORIGIN

Query Match 64.5%; Score 43.2; DB 39; Length 436;
Best Local Similarity 93.8%; Pred. No. 2.2e-05;
Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTTCCATTCGATCAGACATTGTGATGCTGCAACTGGCATAT 57
Db 267 GTTCCATTCGATCAGACATTGTGATGCTGCAACTGGCATAT 220

RESULT 6
A1367541/c 396 bp mRNA EST 15-FEB-1999
LOCUS qv93c12.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:1989142 3',
DEFINITION mRNA sequence.
ACCESSION A1367541
NID 94137286
VERSION A1367541.1 GI:4137286
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 396)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrip/image/image.html

FEATURES
source
Insert Length: 1828 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 370.
Location/Qualifiers
1. 396
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:1989142"
/clone.lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Clone: unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 98 a 86 c 97 g 115 t
ORIGIN

Query Match 64.5%; Score 43.2; DB 45; Length 396;
Best Local Similarity 93.8%; Pred. No. 2.2e-05;
Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTTCCATTCGATCAGACATTGTGATGCTGCAACTGGCATAT 57
Db 278 GTTCCATTCGATCAGACATTGTGATGCTGCAACTGGCATAT 231

RESULT 7
A1555990/c 501 bp mRNA EST 04-MAY-1999
LOCUS t142c03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2243428 3',
DEFINITION mRNA sequence.
ACCESSION A1555990
NID 94739969
VERSION A1555990.1 GI:4739969
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 501)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jun 5, 1998 this sequence version replaced gi:3188057.

JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrip/image/image.html

FEATURES
source
Seq primer: -40UP from GIBCO
High quality sequence stop: 458.
Location/Qualifiers
1. 501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2243428"
/clone.lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:04:19 ; Search time 1962.17 Seconds
(without alignments)
162,081 Million cell updates/sec

Title: US-08-956-518a-99

Perfect score: 100
Sequence: 1 CTGTTTCTAGTGTGCTGATGAG.....CTGCCTCCAGGTAGCTGCA 100

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

ibase :

GenEmbl.*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_bal:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83.4	83.4	1559	9	HUMA7NAR	L25827 Human a7 nt

2	83.4	83.4	2087	10	HSARA7A	X70297 H.sapiens m
3	83.4	83.4	1555	10	HSCHRNA7A	Z23141 H.sapiens C
4	83.4	83.4	1509	10	HSNACHRA7	Y08420 H.sapiens m
5	83.4	83.4	1976	10	HSU62436	U62436 Human nicot
6	82.8	82.8	641	11	AF029838	AF029838 Homo sapi
7	82.8	82.8	689	11	AF029839	AF029839 Homo sapi
8	82.8	82.8	1712	11	AF036903	AF036903 Homo sapi
9	82.8	82.8	1836	11	AF037646	AF037646 Homo sapi
10	81.8	81.8	1977	11	HSU40583	U40583 Human alpha
11	69.0	69.0	1848	12	MOSNARS	L37663 Mus musculu
12	69.0	69.0	2106	12	RATNARAD	L31619 Rattus ratt
13	69.0	69.0	3030	12	RATNARAD	L31619 Rattus ratt
14	67.4	67.4	1551	3	BTAT	S53987 nicotinic r
15	57.8	57.8	2696	4	GGATNAREC	X93604 B.taurus nr
16	57.8	57.8	2037	4	GGNARA7A	X52295 Chicken alp
17	48.8	48.8	2030	4	GGANAREC	X68586 G.gallus nr
18	48.2	48.2	336	14	GA7584	X52296 Chicken alp
19	42.8	42.8	3029	37	AF143847	G47584 Z23357_1 Ze
20	40.0	40.0	89871	37	AC005890	AF13847 Heliothis
21	35.4	35.4	76947	36	AC004326	AC005890 Drosophila
22	33.6	33.6	143664	35	AC007291	AC004326 Drosophila
23	32.6	32.6	1008	12	RATNACHRS	AC007291 Drosophila
24	31.0	31.0	41370	37	CEL30612	M33952 Rat neurona
25	30.6	30.6	2493	12	RNNIACETE	U21319 Caenorabdt
26	30.6	30.6	2461	12	RNT42976	X15834 R.norvegicu
27	30.4	30.4	108487	10	HS738815	U42976 Rattus norv
28	29.6	29.6	2682	3	GOWMTGRG	AL035252 Human DNA
29	29.0	29.0	3629	37	AF143846	M5541 Capra hircu
30	28.8	28.8	1609	11	AF045765	AF13846 Heliothis
31	28.8	28.8	60896	35	AC007582	AF045765 Homo sapi
32	28.2	28.2	3841	4	GGU62904	AC007582 Drosophila
33	28.0	28.0	42782	1	BS43KBDNA	U62904 Gallus gall
34	28.0	28.0	217420	1	BSUB0017	AJ233978 Bacillus
35	28.0	28.0	209510	1	BSUB0018	Z89120 Bacillus su
36	28.0	28.0	4085	4	CHKACHRO2	Z59121 Bacillus su
37	28.0	28.0	415	10	PTU11795	K02904 Chicken nic
38	27.8	27.8	3315	12	MMDMR15A	U11795 Pan troglod
39	27.8	27.8	12371	12	MMDMR15A	Z21504 M.musculus
40	27.6	27.6	85439	37	AC004728	Z38015 M.musculus
41	27.4	27.4	2685	3	BOSMTTGRG	AC004728 Drosophila
42	27.4	27.4	140186	35	AC006404	M6494 Boselapnus
43	27.4	27.4	151528	35	AC006447	AC006404 Mus muscu
44	27.4	27.4	257867	37	AC005557	AC006447 Mus muscu
45	27.0	27.0	339	9	HOMTCRAVE	AC005557 Drosophila

ALIGNMENTS

RESULT 1	HUMA7NAR	1559 bp	MRNA	PRI	18-AUG-1994
LOCUS	Human a7	nicotinic acetylcholine receptor mRNA.			
DEFINITION	L25827				
ACCESSION	G438616				
VERSION	L25827.1	GI:438616			
KEYWORDS	nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.				
SOURCE	Homo sapiens (library: Clontech HL1065b; ATCC 37433) fetus brain CDNA to mRNA.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (sites)				
TITLE	Dovelette-Stamm, L., Monteggia, L.M., Donnelly-Roberts, D., Wang, M.T., Lee, J., Tian, J. and Giordano, T.				
JOURNAL	Cloning and sequence of the human a7 nicotinic acetylcholine receptor				
FEATURES	Drug Dev. Res. (1993) In press				
source	Location/Qualifiers				
	1..1559				
	/Organism="Homo sapiens"				
	/db_xref="taxon:9606"				

/dev_stage="fetus"
/tissue_type="brain"
/tissue_lib="Clontech HL1065D; ATCC 37433"
BASE COUNT 314 a 463 c 440 g 342 t
ORIGIN

Query Match 83.4%; Score 83.4; DB 9; Length 1559;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGGGAGGAGCGGCTTTCAGCCACATTCACACTACGTTGGTGAATCTCTGGG 68
DB 326 AGTGGGAGGAGCGGCTTTCAGCCACATTCACACTACGTTGGTGAATCTCTGGG 68
OY 69 CATTGCCAGTACCTGCTCCAGGTA 93
DB 386 CATTGCCAGTACCTGCTCCAGGCA 410

..SUIT 2

HSARA7A 2087 bp mRNA PRI 01-JUN-1994
LOCUS H.sapiens mRNA for neuronal nicotinic acetylcholine receptor
DEFINITION alpha-7 subunit.
X70297
ACCESSION 9496606
NID X70297.1 GI:496606
VERSION neuronal nicotinic acetylcholine receptor alpha-7 subunit.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2087)
AUTHORS Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.
TITLE Human alpha 7 acetylcholine receptor: cloning of the alpha 7
subunit from the SH-SY5Y cell line and determination of
pharmacological properties of native receptors and functional alpha
7 homomers expressed in Xenopus oocytes
Mol. Pharmacol. 45 (3), 546-554 (1994)
94195283

JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 2087)
AUTHORS Katz, M.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of
Medicine, Dept of Neuroscience, 38th & Hamilton Walk, Rm 235
Stemmler Hall, Philadelphia, PA 19104, USA
LOCATION/Qualifiers
1. .2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/clone_lib="SHSY-5Y"
/clone_lib="lambda zap II"
/clone="SHSY3"
104. .175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
104. .1612
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
/protein_id="CAA49778.1"
/db_xref="PID:9496607"
/db_xref="GI:496607"
/db_xref="SWISS-PROT:P36544"
/translation="MPCSPGVVGLAASLTHVSLQGFQRKLYKELVKNYINLEPRV
ANDSOPLTIVFSLQIMDVDERKNVLTNTILNQSMWDHYLQWVSEPGVKTFR
PDCQIKRFDILYNSADERDARFHTNVLVNSGHCQYLPGLFKSSCYIDVWFPEF
VOCKIKRFGSWSGNSLDLQMGKADISGTFPGEVDLVGIPGRSEERYECCKEKYP
DVFTYTMRRRTIYIGLNLIPCVLLSALALVFLPADSGEKISIGITVLLSLTTFM
LVLAELPATSDSVPLIAQYFASTMIYGVSVVYTVIYQYHHHDPDGGKMRKRWTVI

sig_peptide
CDS

LINMCAMFLRRKRGEDKVRPAQCHKORCLASVENSANVPPASNGNLTYIGRGL
DGVCHVPPDSGVYICGRACCPPTDHEHLHGQPPEDGPDIAKILEEYRIANRRCQ
DESAVSEKFAACVDRCLMAFSVFTIICITIGILMSAPNEVAEVAEKDRA"
BASE COUNT 441 a 601 c 573 g 472 t
ORIGIN

Query Match 83.4%; Score 83.4; DB 10; Length 2087;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGGGAGGAGCGGCTTTCAGCCACATTCACACTACGTTGGTGAATCTCTGGG 68
DB 452 AGTGGGAGGAGCGGCTTTCAGCCACATTCACACTACGTTGGTGAATCTCTGGG 68
OY 69 CATTGCCAGTACCTGCTCCAGGTA 93
DB 512 CATTGCCAGTACCTGCTCCAGGCA 536

RESULT 3

HSCHRNA7A 1555 bp mRNA PRI 03-MAR-1994
LOCUS H.sapiens CHRNA7 mRNA, 3' end.
DEFINITION alpha-7 subunit.
X23141
ACCESSION 9457736
NID X23141.1 GI:457736
VERSION alpha7 nicotinic receptor subunit; cholinergic receptor; CHRNA7
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1555)
AUTHORS Chini, B., Raymond, E., Elgoyhen, A.B., Morali, D., Balzarotti, M. and
Helmenstein, S.
TITLE Molecular cloning and chromosomal localization of the human alpha
7 nicotinic receptor subunit gene (CHRNA7)
Genomics 19 (2), 379-381 (1994)
94245214
REFERENCE 2 (bases 1 to 1555)
AUTHORS Chini, B.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille,
Montpellier, 34094 Cedex, 5, France
LOCATION/Qualifiers
1. .1555
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="retina"
/clone_lib="lambda gt11"
/clone="CHRNA7"
/gene="CHRNA7"
/product="cholinergic receptor, nicotinic, alpha
polypeptide 7"
1. .1440
/gene="CHRNA7"
/gene="CHRNA7"
/codon_start=1
/product="cholinergic receptor, nicotinic, alpha
polypeptide 7"
/protein_id="CAA80672.1"
/db_xref="PID:9457737"
/db_xref="GI:457737"
/db_xref="SWISS-PROT:P36544"
/translation="EQRKLYKELVKNYINLEPRVANDSOPLTIVFSLQIMDVDE
KNQVLTNTILNQSMWDHYLQWVSEPGVKTFRPDCQIKRFDILYNSADERDARF
HTNVLVNSGHCQYLPGLFKSSCYIDVWFPEFVOCKIKRFGSWSGNSLDLQMGK
ADISGTFPGEVDLVGIPGRSEERYECCKEKYPDVFTYTMRRRTIYIGLNLIPCV
LLSALALVFLPADSGEKISIGITVLLSLTTFMVLVAELPATSDSVPLIAQYFAST
MIYGVSVVYTVIYQYHHHDPDGGKMRKRWTVILINMCAMFLRRKRGEDKVRPAQ

mat_peptide
gene
CDS

BASE COUNT
ORIGIN

320 a 454 c 434 g 347 t

OHKORSSLASVENSANVAPPASNGNLLITGFRGLDGVHCYTPDSSGVCAMPAPPT
HDEHLHGOPEDPDIAKILEEVRITANRCDSEAVCSMKRPAACVVDCLIM
AFSVFTICTGILMSAPNFEVANSKDFIA

Query Match 83.4%; Score 83.4; DB 10; Length 1555;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 68
DB 280 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 339
OY 69 CATTGCCAGTACTGCTCCAGGTA 93
DB 340 CATTGCCAGTACTGCTCCAGGTA 364

LOCUS
DEFINITION

HSNACHRA7 1509 bp mRNA PRI 22-JAN-1998
H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit
precursor.

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

gene

mat_peptide

BASE COUNT 297 a 451 c 429 g 332 t

Query Match 83.4%; Score 83.4; DB 10; Length 1509;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 68
DB 349 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 408
OY 69 CATTGCCAGTACTGCTCCAGGTA 93
DB 409 CATTGCCAGTACTGCTCCAGGTA 433

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

LNMCAFLMKRPGEDKYPACQKORRSLASVENSANVAPPASNGNLLITGFRGL
DGVHCYTPDSSGVCAMPAPPT
HDEHLHGOPEDPDIAKILEEVRITANRCDSEAVCSMKRPAACVVDCLIM
AFSVFTICTGILMSAPNFEVANSKDFIA

gene /gene="nAchRA7"
1.1509
67.1506
mat_peptide /gene="nAchRA7"

Query Match 83.4%; Score 83.4; DB 10; Length 1509;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 68
DB 349 AGTGTGATGAGCGCTTTGAGCCGACATTCACACTAAGTGTGTGATCTCTGGG 408
OY 69 CATTGCCAGTACTGCTCCAGGTA 93
DB 409 CATTGCCAGTACTGCTCCAGGTA 433

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

mat_peptide

LVAEIPEATSDVPELIAQFEASTMTIVGLSVVTVIVIQYHHHDPDGKMKPRTRVI
LWMCAMFELMKRPPGDKVAPACOKRORSLASVEMSAVAPPANGLNLTIGFGL
DGVHCVPTDPSGVCGRMACSPHDEHLHGGPPGDDPLAKILEEVXIRYNRFQCO
DESEAVCSSEKRFACVYDRCLMAFVFTIICIGILMARPNVEVNSDFA
139, 1578
/product="nicotinic acetylcholine receptor alpha7 subunit"

3'UTR
BASE COUNT 369 a 533 c 531 g 423 t
ORIGIN

Query Match 83.4%; Score 83.4; DB 10; Length 1876;
Best Local Similarity 98.8%; Pred. No. 1.6e-20;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 AGTGTGATGAGCGCTTGACGCACATTCACACTAGCTGTGGATTTCTTGGG 68
421 AGTGTGATGAGCGCTTGACGCACATTCACACTAGCTGTGGATTTCTTGGG 480
VY 69 CATGGCAGTACTGCTCCAGGTA 93
DB 481 CATGGCAGTACTGCTCCAGGCA 505

RESULT 6
AF029838 641 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION
ACCESSION AF029838
NID 93757793
VERSION AF029838.1 GI:3757793
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 641)
AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J., Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K., Bulding,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.,
Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 641)
AUTHORS Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Breese,C., Davis,A., Hopkins,J. and Freedman,R.,
Direct Submission
JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES
SOURCE Location/Qualifiers
1. 641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR1"
/note="RACE product A/B/C/D"
1. >641
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

BASE COUNT 151 a 192 c 163 g 135 t
ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 641;
Best Local Similarity 97.7%; Pred. No. 2.4e-20;
Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

3ATGAGCGCTTGAGCGCATTCACACTAGCTGTGGATTTCTTGG 67
TACGCGCTTGAGCGCATTCACACTAGCTGTGGATTTCTTGG 591

QY 68 GCATTGCCAGTACTGCTCCAGGTA 93
DB 592 GCATTGCCAGTACTGCTCCAGGCA 617

RESULT 7
AF029839 689 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION
ACCESSION AF029839
NID 93757794
VERSION AF029839.1 GI:3757794
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 689)
AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J., Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K., Bulding,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.,
Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 689)
AUTHORS Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Breese,C., Davis,A., Hopkins,J. and Freedman,R.,
Direct Submission
JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES
SOURCE Location/Qualifiers
1. 689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
1. >689
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

BASE COUNT 142 a 210 c 196 g 141 t
ORIGIN

Query Match 82.8%; Score 82.8; DB 11; Length 689;
Best Local Similarity 97.7%; Pred. No. 2.4e-20;
Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TAGTGTGATGAGCGCTTGAGCGCATTCACACTAGCTGTGGATTTCTTGG 67
DB 580 TAGTGTGATGAGCGCTTGAGCGCATTCACACTAGCTGTGGATTTCTTGG 639
QY 68 GCATTGCCAGTACTGCTCCAGGTA 93
DB 640 GCATTGCCAGTACTGCTCCAGGCA 665

RESULT 8
AF036903 1712 bp mRNA PRI 16-OCT-1998
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION
ACCESSION AF036903
NID 93757807
VERSION AF036903.1 GI:3757807
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1712)

FEATURES	source
1. 1836	/organism="Homo sapiens"
2 (bases 1 to 1836)	Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Breese,C., Davis,A., Hopkins,J., and Freedman,R.
3	Direct Submission
4	Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
5	Location/Qualifiers
6	Location/Qualifiers
7	Location/Qualifiers
8	Location/Qualifiers
9	Location/Qualifiers
10	Location/Qualifiers
11	Location/Qualifiers
12	Location/Qualifiers
13	Location/Qualifiers
14	Location/Qualifiers
15	Location/Qualifiers
16	Location/Qualifiers
17	Location/Qualifiers
18	Location/Qualifiers
19	Location/Qualifiers
20	Location/Qualifiers
21	Location/Qualifiers
22	Location/Qualifiers
23	Location/Qualifiers
24	Location/Qualifiers
25	Location/Qualifiers
26	Location/Qualifiers
27	Location/Qualifiers
28	Location/Qualifiers
29	Location/Qualifiers
30	Location/Qualifiers
31	Location/Qualifiers
32	Location/Qualifiers
33	Location/Qualifiers
34	Location/Qualifiers
35	Location/Qualifiers
36	Location/Qualifiers
37	Location/Qualifiers
38	Location/Qualifiers
39	Location/Qualifiers
40	Location/Qualifiers
41	Location/Qualifiers
42	Location/Qualifiers
43	Location/Qualifiers
44	Location/Qualifiers
45	Location/Qualifiers
46	Location/Qualifiers
47	Location/Qualifiers
48	Location/Qualifiers
49	Location/Qualifiers
50	Location/Qualifiers
51	Location/Qualifiers
52	Location/Qualifiers
53	Location/Qualifiers
54	Location/Qualifiers
55	Location/Qualifiers
56	Location/Qualifiers
57	Location/Qualifiers
58	Location/Qualifiers
59	Location/Qualifiers
60	Location/Qualifiers
61	Location/Qualifiers
62	Location/Qualifiers
63	Location/Qualifiers
64	Location/Qualifiers
65	Location/Qualifiers
66	Location/Qualifiers
67	Location/Qualifiers
68	Location/Qualifiers
69	Location/Qualifiers
70	Location/Qualifiers
71	Location/Qualifiers
72	Location/Qualifiers
73	Location/Qualifiers
74	Location/Qualifiers
75	Location/Qualifiers
76	Location/Qualifiers
77	Location/Qualifiers
78	Location/Qualifiers
79	Location/Qualifiers
80	Location/Qualifiers
81	Location/Qualifiers
82	Location/Qualifiers
83	Location/Qualifiers
84	Location/Qualifiers
85	Location/Qualifiers
86	Location/Qualifiers
87	Location/Qualifiers
88	Location/Qualifiers
89	Location/Qualifiers
90	Location/Qualifiers
91	Location/Qualifiers
92	Location/Qualifiers
93	Location/Qualifiers
94	Location/Qualifiers
95	Location/Qualifiers
96	Location/Qualifiers
97	Location/Qualifiers
98	Location/Qualifiers
99	Location/Qualifiers
100	Location/Qualifiers

PDGQIMKPDILLVNSADERPDATFHTNVLNPSGHCOYLPGIKRSEFCYIDVREPPD
VOHCKIKFGSMYSYGLSLDLOMGEADISGYPINPGEHMDVIGIPGRSEFFECCEPFB
DVTETVMTARRRTLYGMLLIPCVILSLALVLELPDSEKISLGTIVLLSTVEM
LIVAEIMPATSDSVPLIAOYFASMTIIVGLSVVTVIYLYQHHNDPDGKMPKTRVLI
LLNWCAMFLMRKRGEDVVRPACQKORCSLASVEMSAVAPPAASGNLLYIGFRL
DGVHCVPTPDGSGVRMACSPTEHDEHLHGGOPEPBGPDIAKILEEVRVIANFRCO
DESEVICSEMKRACVDRCLMAFSVFTIICITIGILMSAPNFEAVSKDPA"

3'UTR
polyA_site 1514..1977
BASE COUNT 426 a 567 c 524 g 460 t
ORIGIN

Query Match 81.8% Score 81.8: DB 10: Length 1977;
Best Local Similarity 97.6% Pred. No. 6.2e-20;
Matches 83: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTGTGATGATCTCTGGG 68
356 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTGTGATGATCTCTGGG 415
|||||
QY 69 CATGCCAGTACCTGCTCCAGGTA 93
|||||
DB 416 CATGCCAGTACCTGCTCCAGGTA 440

RESULT 11
MUSNARS 1848 bp mRNA ROD 02-FEB-1999
LOCUS Mus musculus neuronal nicotinic acetylcholine receptor subunit
DEFINITION alpha 7 mRNA, complete cds.
ACCESSION L37663.1
NID L37663.1 GI:790853
VERSION L37663.1
KEYWORDS neuronal acetylcholine receptor subunit alpha 7 unit;
SOURCE Mus musculus (strain BALB/c, sub-species domesticus) neonatal brain
cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1848)
Orr-Urtreger, A., Seldin, M.F., Baldini, A. and Beaudet, A.L.
Cloning and mapping of the mouse alpha 7-neuronal nicotinic
acetylcholine receptor
JOURNAL Genomics 26 (2), 399-402 (1995)
MEDLINE 95324936

JOURNAL
MEDLINE
SOURCE

1. 1848
/organism="Mus musculus"
/strain="BALB/c"
/sub-species="domesticus"
/db_xref="taxon:10090"
/dev_stage="neonatal"
/tissue_type="brain"
1..50
51..1559
/codon_start=1
/product="neuronal acetylcholine receptor subunit alpha 7"
/protein_id="AAC42053.1"
/db_xref="PID:g790854"
/db_xref="GI:790854"
/translation="MCGRRGIVLAAALHVSLOGEFORRLYELVKNYNPRLERPV
ANDSPLTYFSLSLQINDVDEKQVLTINIMLQMSWTDHYLQNMNSEYGVKNVRF
PQGLKRPDILLNSADERDATERHTNVLNPSGHCOYLPGIKRSEFCYIDVREPPD
VOGCKIKFGSMYSYGLSLDLOMGEADISYIPNGEMDLMGIPGRKNEKFECCKEPYP
DVTETVMTARRRTLYGMLLIPCVILSLALVLELPDSEKISLGTIVLLSTVEM
LIVAEIMPATSDSVPLIAOYFASMTIIVGLSVVTVIYLYQHHNDPDGKMPKTRVLI
LLNWCAMFLMRKRGEDVVRPACQKORCSLASVEMSAVAPPAASGNLLYIGFRL
BGMHCAPTPDGSVVGGRGLACSPTEHDEHLHGHPDGDPLAKILEEVRVIANFRCO
DESEVICSEMKRACVDRCLMAFSVFTIICITIGILMSAPNFEAVSKDPA"

3'UTR
BASE COUNT 406 a 487 c 500 g 455 t
ORIGIN

Query Match 69.0% Score 69: DB 12: Length 1848;
Best Local Similarity 88.2% Pred. No. 3.1e-15;
Matches 75: Conservative 0; Mismatches 10; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTGTGATGATCTCTGGG 68
399 AGTGTGATGAGCGCTTGACGCCACATTCACACTAGCTGTGATGATCTCTGGG 458
|||||
QY 69 CATGCCAGTACCTGCTCCAGGTA 93
|||||
DB 459 CATGCCAGTACCTGCTCCAGGTA 483

RESULT 12
RATNARAD 2106 bp mRNA ROD 28-DEC-1998
LOCUS Rattus rattus nicotinic acetylcholine receptor alpha 7 subunit
DEFINITION mRNA, complete cds.
ACCESSION L31619.1
NID L31619.1 GI:3478618
VERSION L31619.1
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 2106)
Boulter, J.
TITLE Nicotinic acetylcholine receptor subunit gene alpha 7
JOURNAL Unpublished (1994)
AUTHORS Boulter, J.
REFERENCE 2 (bases 1 to 2106)
JOURNAL Unpublished (1992)
AUTHORS Boulter, J.
REFERENCE 3 (bases 1 to 2106)
JOURNAL Direct Submission
TITLE Submitted (06-APR-1994) Department of Psychiatry and Biobehavioral
Sciences, University of California, 760 Westwood Plaza, Los
Angeles, CA 90095-1759, USA
4 (bases 1 to 2106)
Hartley, M.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) The Salk Institute, MNJ-H, 10010 North
Torrey Pines Road, La Jolla, CA 92037, USA
Sequence update by submitter
On Aug 27, 1998 this sequence version replaced gi:468919.

REMARK
COMMENT
FEATURES
SOURCE

1. 2106
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_type="neuronal"
/dev_stage="adult"
/tissue_type="brain"
23..1531
/codon_start=1
/product="nicotinic acetylcholine receptor alpha 7
subunit"
/protein_id="AAC33136.1"
/db_xref="PID:g3478619"
/db_xref="GI:3478619"
/translation="MCGRRGIVLAAALHVSLOGEFORRLYELVKNYNPRLERPV
ANDSPLTYFSLSLQINDVDEKQVLTINIMLQMSWTDHYLQNMNSEYGVKNVRF
PQGLKRPDILLNSADERDATERHTNVLNPSGHCOYLPGIKRSEFCYIDVREPPD
VOGCKIKFGSMYSYGLSLDLOMGEADISYIPNGEMDLMGIPGRKNEKFECCKEPYP
DVTETVMTARRRTLYGMLLIPCVILSLALVLELPDSEKISLGTIVLLSTVEM
LIVAEIMPATSDSVPLIAOYFASMTIIVGLSVVTVIYLYQHHNDPDGKMPKTRVLI
LLNWCAMFLMRKRGEDVVRPACQKORCSLASVEMSAVAPPAASGNLLYIGFRL
BGMHCAPTPDGSVVGGRGLACSPTEHDEHLHGHPDGDPLAKILEEVRVIANFRCO
DESEVICSEMKRACVDRCLMAFSVFTIICITIGILMSAPNFEAVSKDPA"

BASE COUNT 473 a 554 c 533 g 546 t

ORIGIN

Query Match 69.0%; Score 69; DB 12; Length 2106;
 Best Local Similarity 88.2%; Pred. No. 3.2e-15;
 Matches 75; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 AGTCGTATGAGCGCTTGACGCCACATTCACACTAAGCTGGTGGATCTCTGGG 68
 |||
 DB 371 AGTCGTATGAGCGCTTGATGATGACACATGTTTGGTGAATGATCTGGG 430
 |||
 QY 69 CATGCCAGTACCTGCTCCAGGTA 93
 |||
 DB 431 CATGCCAGTACCTGCTCCAGGTA 455

RESULT 13

LOCUS S53987 3030 bp mRNA ROD 19-MAR-1993
 INITIATION nicotinic receptor alpha 7 subunit [rats, brain, mRNA, 3030 nt].
 ESSION S53987
 NID Q264770
 VERSION S53987.1 GI:264770
 KEYWORDS
 SOURCE Rattus sp. brain.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 3030)
 Seguela, P., Wadiche, J., Dineley-Keller, K., Dani, J. A. and
 Patrick, J. W.
 Molecular cloning, functional properties, and distribution of rat
 brain alpha 7: a nicotinic cation channel highly permeable to
 calcium
 J. Neurosci. 13 (2), 596-604 (1993)

JOURNAL MEDLINE 93147931
 REMARK GenBank staff at the National Library of Medicine created this
 entry [NCBI g1dbseq 124020] from the original journal article.
 This sequence comes from Fig. 1.

FEATURES

source
 1. 3030
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 42. 1550
 /gene="nicotinic receptor alpha 7 subunit"
 42. 1550
 /note="This sequence comes from Fig. 1. Protein sequence
 is in conflict with the conceptual translation;
 mismatch(469[R->P])"
 /codon_start=1
 /product="nicotinic receptor alpha 7 subunit"
 /protein_id="AA825224.1"
 /db_xref="PID:g264771"
 /db_xref="GI:264771"
 /translation="MCGGKGIGLALAAALLHVSLOGEFORRLKYLKYNPLERPV
 ANDSPATVYFSLQINDVDEKNOVLTNTIMLQMSWDHYLQNNASEYGVKRVF
 PDSOIMKPDILLYNSADERDATFTNVLVNSGHCYLPPIFKSSCYIDVPMFPDQ
 VQCKIKFGSSYGGSLDLOMOEADISYINGEMDLVGVGRSEKREYECCKEYPPV
 DVVYTTMRRRTLYGNNLILICVLSALALVFLPADSGKISLGLTVLSIVFM
 LVAELIPATSDSVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGMKRWTRIL
 LKWCAMFLMKRFGEDKVRPAQOHKRRCSLASVMSVAPPTNGNLILIGRGLDTM
 EGNHCAPTSDSGVGRVACSPTHDEHLHGAQPSGPDLAKEVRYIAHRPDCODES
 DESEVISEKFAACVDRCLMAFVSFTILCTIGLMSAPNFEVAVSKDFA"

BASE COUNT 698 a 756 c 801 g 775 t
 ORIGIN

Query Match 69.0%; Score 69; DB 12; Length 3030;
 Best Local Similarity 88.2%; Pred. No. 3.2e-15;
 Matches 75; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 AGTCGTATGAGCGCTTGACGCCACATTCACACTAAGCTGGTGGATCTCTGGG 68

DB 390 AGTCGTATGAGCGCTTGATGACGCCACATTCACACTAAGCTGGTGGATCTCTGGG 449
 |||
 QY 69 CATGCCAGTACCTGCTCCAGGTA 93
 |||
 DB 450 CATGCCAGTACCTGCTCCAGGTA 474

RESULT 14

LOCUS BTA7 1551 bp mRNA MAM 09-JAN-1996
 DEFINITION B. taurus mRNA for alpha7 nicotinic acetylcholine receptor subunit.
 ACCESSION X93604
 NID 91181202
 VERSION X93604.1 GI:1181202
 KEYWORDS alpha7 gene; alpha7 nicotinic receptor; alpha7 subunit; nicotinic
 acetylcholine receptor.
 SOURCE Bos taurus.
 ORGANISM Bos taurus.
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 1551)
 Garcia-Guzman, M., Sala, F., Sala, S., Campos-Caro, A., Stuhmer, W.,
 Gutierrez, L. M. and Criado, M.
 alpha-Bungarotoxin-sensitive nicotinic receptors on bovine
 chromaffin cells: molecular cloning, functional expression and
 alternative splicing of the alpha 7 subunit
 Eur. J. Neurosci. 7 (4), 647-655 (1995)

JOURNAL MEDLINE 95346009
 REMARK 2 (bases 1 to 1551)
 AUTHORS Garcia-Guzman, M.
 TITLE Direct Submission
 JOURNAL Submitted (30-NOV-1995) M. Garcia-Guzman, Max-Planck Institut fuer
 FUEER EXPERIMENTELLE MEDIZIN (Abt. XI), Hermann-Rein strasse, 3, D-
 37075-Goettingen, FRG
 Revised by [3]
 COMMENT On Feb 6, 1996 this sequence version replaced g1:1103382.
 FEATURES
 source
 1. 1551
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /dev_stage="adult"
 /tissue_type="adrenal gland medulla"
 43. 99
 /gene="alpha7"
 43. 1542
 /gene="alpha7"
 /codon_start=1
 /product="alpha7 nicotinic acetylcholine receptor subunit"
 /protein_id="CAA63802.1"
 /db_xref="PID:g212821"
 /db_xref="PID:g1103383"
 /db_xref="GI:1103383"
 /db_xref="SWISS-PROT:P54131"
 /translation="MRGSLCLALASILHVSLOGEFORRLKYLKYNPLERPVAND
 SPLVYFSLQINDVDEKNOVLTNTIMLQMSWDHYLQNNASEYGVKRVFDPDQ
 QVCKIKFGSSYGGSLDLOMOEADISYINGEMDLVGVGRSEKREYECCKEYPPV
 DVVYTTMRRRTLYGNNLILICVLSALALVFLPADSGKISLGLTVLSIVFM
 LVAELIPATSDSVPLIAQYFASMTIVGLSVVTVIVLQHHDPDGGMKRWTRIL
 LKWCAMFLMKRFGEDKVRPAQOHKRRCSLASVMSVAPPTNGNLILIGRGLDTM
 HCAPPTSDSGVGRVACSPTHDEHLHGAQPSGPDLAKEVRYIAHRPDCODES
 EAVCSWKRFAACVDRCLMAFVSFTILCTIGLMSAPNFEVAVSKDFA"

CDS sig-peptide

BASE COUNT 305 a 477 c 429 g 340 t
 ORIGIN
 mat-peptide
 /gene="alpha7"
 100. 1539
 /gene="alpha7"

Query Match 67.4%; Score 67.4; DB 3; Length 1551;
 Best Local Similarity 87.1%; Pred. No. 1.2e-14;

	Matches	74;	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
OY	9	AGTGTGATGACGGTTTGACGCACATTCCACACTAAGCTTGGTAATCTTTGGG	68							
Dδ	382	AGTGTGATGACGGTTTGATCTCACTCAGTCCTCCACACCAGAAGTGTATGATAATCTTTGGG	441							
OY	69	CATGCCAGTACTCTCGCTCCACAGTA	93							
Dδ	442	CAGTCGCAGTACTCTCGCTCCACAGGA	466							

	Matches	68;	Conservative	0;	Mismatches	17;	Indels	0;	Gaps	0;
QY	9	AGCTGCTATGAGCGCTTGAGCCGACATTCACACTAACGCTGGTGAATCTTCTGGG	68							
Db	353	AGCTGCTATGAGCGCTTGAGCTTCACTTTCACACTTAAGTTTAGTCATCTTGGGA	412							
QY	69	CATTGCCAGTACTGCTCCACAGTA	93							
Db	413	CACGTGCCAATATCTGCCACACAGGCA	437							

Search completed: September 17, 1999, 22:04:21
Job time: 16444 sec

```

RESULT 15
GGATNAREC
LOCUS      GGATNAREC      2696 bp      mRNA      VRT      12-AUG-1996
DEFINITION Chicken alpha7 subunit of nicotinic acetylcholine receptor.
ACCESSION  X52295
NID         963077
VERSION    X52295.1 GI:63077
WORDS      alpha7 subunit; nicotinic acetylcholine receptor.
'RCB
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
            Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 2696)
AUTHORS   Schoeffer,R.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-1990) Schoeffer R., The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
REFERENCE  2 (bases 1 to 2696)
AUTHORS   Schoeffer,R., Conroy,W.G., Whitting,P., Gore,M. and Lindstrom,J.
TITLE     Brain alpha-bungarotoxin binding protein cDNAs and MAbS reveal
            subtypes of this branch of the ligand-gated ion channel gene
            superfamily
JOURNAL   Neuron 5 (1), 35-48 (1990)
MEDLINE   90315158
COMMENT   See also <X52296>.
FEATURES   Location/Qualifiers
            1..2696

```

sig_peptide
CDS
nat_peptide

```
1 /clone="pCh29-3 and pCh34-1."  
2 5...70  
3 /product="signal peptide (AA -22 to -1)"  
4 5...1513  
5 /note="protein precursor (AA -21-480)"  
6 /codon_start=1  
7 /product="alpha7 subunit of nicotinic acetylcholine  
8 receptor"  
9 /protein_id="CAA35543.1"  
10 /db_xref="pid:963078"  
11 /db_xref="gi:63078"  
12 /translation="MGLRALMLLAAAGLVRESIQGEFQRLKELIKLKNYPLERPV  
13 ANDSPLEVFRLSLMDIQMDENKQVLTITNIMQMTDHYLQMNYSKGVKVFPR  
14 PGDLIRGMDILLYNADERPDATPEFTENILNVNSGCDPLGKIRSCYIDYRNPFD  
15 VOKMLKRSMTSGMSTSLDMDQNDISGYISNSGMDLVGIPGRKTESFTCCKEPFR  
16 DITPEYRRRRRLIYGLMLLPCVLISLALVLPDSEKISLGLTVLSTVEM  
17 LLVAIPATISVPLINAOIFRSTVITGLSVVTVLYQYHHHDPOGGKPKRTVYD  
18 LLMNQALRLMKRPEEDKVRPACQKHQRKRSKSNKEMVSGQCSNGNMLYIGFRLI  
19 DGHVQTPPTDSDVIGKMTGSPTEENILHSGHSEGGPDLAKILIEERYIYANFPRD  
20 DEEENIKENMFRAASVDRCLIMAFVETIICTIGILMSAPNEAVASKDA"  
21 .15130
```

BASE COUNT	777 a	/product="mature protein (AA 1-480),"	580 c	555 g	784 t
ORIGIN					
Query Match			57.88;	Score 57.8;	DB 4;
Best Local Similarity			80.08;	Pred. No. 4.3e-11;	Length 2696

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:34:33 ; Search time 425.19 Seconds

(without alignments)
58.842 Million cell updates/sec

Title: us-08-956-518a-99

Perfect score: 100

Sequence: 1 CTGTTCTAGTGTGATGAG.....CTGCCCTCCAGGTAAGTCA 100

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.4	83.4	1876	1 T48239	Neuronal nicotinic
2	83.4	83.4	1876	1 V12197	Human neuronal nic
3	83.4	83.4	1590	1 V44687	V2747 variant huma
4	57.8	57.8	2769	1 T59196	Neuronal alpha-bun
5	48.8	48.8	2101	1 T59197	Neuronal alpha-bun
6	30.6	30.6	2460	1 Q06086	Plasmid p2PC13 enc
7	28.4	28.4	2192	1 X23530	Matz x21 gene DT
8	26.6	26.6	1365	1 T79976	Human galanin rece
9	26.6	26.6	1083	1 V10329	Human GalR2 recept
10	26.6	26.6	3390	1 V44931	Human galanin rece
11	26.6	26.6	3390	1 V28291	Human galanin rece
12	26.6	26.6	3390	1 V22649	Human galanin rece
13	25.2	25.2	13574	1 X13051	Enterococcus faeca
14	24.8	24.8	5002	1 X25011	Anther-specific ES
15	24.4	24.4	2376	1 X16153	Mouse Sox1 cDNA. I
16	24.4	24.4	443	1 X21047	Polynucleotide seq
17	24.2	24.2	1934	1 Q91198	HMG 46 kDa antige
18	24.2	24.2	1384	1 Q91199	HMG 46 kDa antige
19	24.0	24.0	1734	1 T61334	Batten disease gen
20	24.0	24.0	1732	1 T61335	Batten disease gen
21	24.0	24.0	1732	1 T61336	Batten disease gen
22	24.0	24.0	1732	1 T61337	Batten disease gen
23	24.0	24.0	1732	1 T61338	Batten disease gen
24	24.0	24.0	1732	1 T61339	Batten disease gen
25	24.0	24.0	1732	1 T61340	Batten disease gen
26	24.0	24.0	1732	1 T61341	Batten disease gen
27	24.0	24.0	1732	1 T61342	Batten disease gen
28	24.0	24.0	1732	1 T61343	Batten disease gen
29	24.0	24.0	1732	1 T61344	Batten disease gen
30	24.0	24.0	1732	1 T61345	Batten disease gen
31	24.0	24.0	1732	1 T61346	Batten disease gen
32	24.0	24.0	1732	1 T61347	Batten disease gen
33	24.0	24.0	1732	1 T61348	Batten disease gen
34	24.0	24.0	1732	1 T61332	Batten disease gen
35	24.0	24.0	1732	1 T61333	Batten disease gen
36	24.0	24.0	1732	1 T61306	Batten disease gen
37	24.0	24.0	16636	1 X20546	Polynucleotide seq
38	23.8	23.8	2038	1 Q89844	Human death associ
39	23.6	23.6	11601	1 V60292	DNA sequence encod
40	23.6	23.6	13058	1 O13608	ACV synthetase gen
41	23.6	23.6	13058	1 O48231	Vector containing
42	23.6	23.6	13058	1 V52256	Streptococcus pneu
43	23.6	23.6	1569	1 V59650	Human secreted pro

44 23.4 23.4 3039 1 Q91096 Human herpesvirus-
c 45 23.4 23.4 8738 1 T72327 Lactobacillus bact

ALIGNMENTS

RESULT 1
ID T48239 standard; DNA; 1876 BP.
AC T48239;
DT 09-APR-1997 (first entry)
DE Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.
KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;
RW ligand-gated receptor; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 73..1581
FT /*tag= a
FN W09641876-A1.
PD 27-DEC-1996.
PE 07-JUN-1996; U09775.
PR 07-JUN-1995; US-484722.
PA (SIBT-) SIBT NEUROSCIENCES INC.
PI Eli Lilly & Co., Indianapolis, IN.
DR P-PSDB; W09025.
DR P-PSDB; W09025.
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
used in screening to determine the effect of drugs on the receptor
PS Disclosure; Page 71-73; 108pp; English.
CC A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of
the human neuronal nicotinic acetylcholine receptor (nAChR). Host
cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7
nucleic acids, opt. in combination with other alpha and/or beta
subunit nucleic acids (see also T48232-38, T48240-41), express
recombinant nAChR subunits useful for identifying cpds. that
modulate the activity of human nAChRs.
CC Sequence 1876 BP; 369 A; 531 G; 423 T;
SQ

Query Match 83.4%; Score 83.4; DB 1; Length 1876;
Best Local Similarity 98.8%; Pred. No. 5.6e-22;
Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 AGTGTGATGAGCGCTTGAGCCACATTCACACTAAGTGTGATCTTCGG 68
DB 421 AGTGTGATGAGCGCTTGAGCCACATTCACACTAAGTGTGATCTTCGG 480
OY 69 CATGCCAGTACCTGCTCCAGGTA 93
DB 481 CATGCCAGTACCTGCTCCAGGCA 505

RESULT 2
ID V12197 standard; cDNA; 1876 BP.
AC V12197;
DT 14-MAY-1998 (first entry)
DE Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.
KW Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;
KW brain tissue; screening; NACNR; antibody; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 73..1581
FT /*tag= a
FT /product= "neuronal nicotinic acetylcholine receptor
alpha-7 subunit".
FN W09420617-A2.
PD 15-SEP-1994.
PE 08-MAR-1994; U02447.
PR 08-MAR-1993; US-028031.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PA (SIBT-) SIBT NEUROSCIENCES INC.


```

DB 419 AGTGGCAGTGAAGATTGATGCTACATTTCACACATGATTAGTACATTTTGGGA 478
OY 69 CATGCCAGTACCTGCTCCAGSTA 93
DB 479 CACTGCCAATATCTGCCACAGSCA 503

```

RESULT 5

```

TS9197
ID TS9197 standard; cDNA: 2101 BP.
AC TS9197;
DT 17-JUN-1997 (first entry)
DE Neuronal alpha-bungarotoxin binding protein alpha2 subunit cDNA.
KW Neuronal alpha-bungarotoxin binding protein alpha 2; cholinergic;
KM ligand binding; ion channel; ss.
OS Gallus sp.
FH Key Location/Qualifiers
cds 56..1501
    signal_peptide 56..145
    mat_peptide 146..1498
    /tag= a
    /tag= b
    /tag= c
    /tag= c
US5599709-A.
PD 04-FEB-1997.
PF 28-SEP-1989; US-413947.
PR 28-SEP-1989; US-413947.
PI (SALK ) SALK INST BIOLOGICAL STUDIES.
PI Lindstrom JM, Schoepfer RD;
DR MPI; 97-118297/11.
P-PSDB; W12369.
PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
PT to screen cholinergic agents and other drugs which may affect ligand
PT binding; ion channel or other activities of the protein.
PS Claim 1; Fig 3A-B; 18pp; English.
CC 2 cDNA clones (TS9196 and TS9197) respectively code for the alpha1
CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
CC an 18-day embryo chick brain cDNA library using a probe (see also
CC TS9198) based on the N-terminal amino acid sequence of chick
CC brain ABBP. The probe isolated partial clone pCH29-1, which
CC encoded the N-terminal portion of alpha1. A subclone, pCH29-3
CC (ATCC 40641), was used to rescreen the library, yielding clone
CC pCH31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
CC A probe based on the C-terminal region of pCH31-1 was used obtain
CC clone pCH34-1 (ATCC 40639), encoding the C-terminal portion of
CC alpha1. The cDNA clones can be used as probes to identify further
CC ABBP subunits, and in the recombinant prodn. of ABBP.
CU Sequence 2101 BP; 582 A; 406 C; 454 G; 659 T;
SQ

```

```

Query Match 48.8%; Score 48.8; DB 1; Length 2101;
Best Local Similarity 70.7%; Pred. No. 5.2e-09;
Matches 65; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

```

OY 9 AGTGGTATGAGCGCTTGAGCGCCATTCACACTAAGCTGGTGGTAATCTCTCGG 68
DB 428 AGTGGGATGAAGATTGATGCAACATTTCACACAATGCTGGTGAATTAATCTGGA 487
OY 69 CATGCCAGTACCTGCTCCAGSTAAGCTGCA 100
DB 488 TCCTGTCAATATATCTCTCCAGGCAATTTGAA 519

```

RESULT 6

```

006086
ID 006086 standard; cDNA: 2460 BP.
AC 006086;
DT 24-JAN-1991 (first entry)
DE Plasmid pPC13 encoding neuronal nicotinic acetylcholine receptor
DE beta 4 subunit.
KW Rat; nAChR; ss.

```

```

OS Rattus rattus.
FH Key Location/Qualifiers
FH cds 61..1548
    signal_peptide 61..120
    mat_peptide 121..1545
    /tag= a
    /tag= b
    /tag= c
    /product=nAChR beta 4 subunit
    /tag= d
    /tag= e
    /tag= f
    /tag= g
    /tag= h
    /tag= i
    /tag= j
    /tag= k
    /tag= l
    /tag= m
    /tag= n
    /tag= o
    /tag= p
    /tag= q
    /tag= r
    /tag= s
    /tag= t
    /tag= u
    /tag= v
    /tag= w
    /tag= x
    /tag= y
    /tag= z
    /tag= aa
    /tag= ab
    /tag= ac
    /tag= ad
    /tag= ae
    /tag= af
    /tag= ag
    /tag= ah
    /tag= ai
    /tag= aj
    /tag= ak
    /tag= al
    /tag= am
    /tag= an
    /tag= ao
    /tag= ap
    /tag= aq
    /tag= ar
    /tag= as
    /tag= at
    /tag= au
    /tag= av
    /tag= aw
    /tag= ax
    /tag= ay
    /tag= az
    /tag= ba
    /tag= bb
    /tag= bc
    /tag= bd
    /tag= be
    /tag= bf
    /tag= bg
    /tag= bh
    /tag= bi
    /tag= bj
    /tag= bk
    /tag= bl
    /tag= bm
    /tag= bn
    /tag= bo
    /tag= bp
    /tag= bq
    /tag= br
    /tag= bs
    /tag= bt
    /tag= bu
    /tag= bv
    /tag= bw
    /tag= bx
    /tag= by
    /tag= bz
    /tag= ca
    /tag= cb
    /tag= cc
    /tag= cd
    /tag= ce
    /tag= cf
    /tag= cg
    /tag= ch
    /tag= ci
    /tag= cj
    /tag= ck
    /tag= cl
    /tag= cm
    /tag= cn
    /tag= co
    /tag= cp
    /tag= cq
    /tag= cr
    /tag= cs
    /tag= ct
    /tag= cu
    /tag= cv
    /tag= cw
    /tag= cx
    /tag= cy
    /tag= cz
    /tag= da
    /tag= db
    /tag= dc
    /tag= dd
    /tag= de
    /tag= df
    /tag= dg
    /tag= dh
    /tag= di
    /tag= dj
    /tag= dk
    /tag= dl
    /tag= dm
    /tag= dn
    /tag= do
    /tag= dp
    /tag= dq
    /tag= dr
    /tag= ds
    /tag= dt
    /tag= du
    /tag= dv
    /tag= dw
    /tag= dx
    /tag= dy
    /tag= dz
    /tag= ea
    /tag= eb
    /tag= ec
    /tag= ed
    /tag= ee
    /tag= ef
    /tag= eg
    /tag= eh
    /tag= ei
    /tag= ej
    /tag= ek
    /tag= el
    /tag= em
    /tag= en
    /tag= eo
    /tag= ep
    /tag= eq
    /tag= er
    /tag= es
    /tag= et
    /tag= eu
    /tag= ev
    /tag= ew
    /tag= ex
    /tag= ey
    /tag= ez
    /tag= fa
    /tag= fb
    /tag= fc
    /tag= fd
    /tag= fe
    /tag= ff
    /tag= fg
    /tag= fh
    /tag= fi
    /tag= fj
    /tag= fk
    /tag= fl
    /tag= fm
    /tag= fn
    /tag= fo
    /tag= fp
    /tag= fq
    /tag= fr
    /tag= fs
    /tag= ft
    /tag= fu
    /tag= fv
    /tag= fw
    /tag= fx
    /tag= fy
    /tag= fz
    /tag= ga
    /tag= gb
    /tag= gc
    /tag= gd
    /tag= ge
    /tag= gf
    /tag= gh
    /tag= gi
    /tag= gj
    /tag= gk
    /tag= gl
    /tag= gm
    /tag= gn
    /tag= go
    /tag= gp
    /tag= gq
    /tag= gr
    /tag= gs
    /tag= gt
    /tag= gu
    /tag= gv
    /tag= gw
    /tag= gx
    /tag= gy
    /tag= gz
    /tag= ha
    /tag= hb
    /tag= hc
    /tag= hd
    /tag= he
    /tag= hf
    /tag= hg
    /tag= hh
    /tag= hi
    /tag= hj
    /tag= hk
    /tag= hl
    /tag= hm
    /tag= hn
    /tag= ho
    /tag= hp
    /tag= hq
    /tag= hr
    /tag= hs
    /tag= ht
    /tag= hu
    /tag= hv
    /tag= hw
    /tag= hx
    /tag= hy
    /tag= hz
    /tag= ia
    /tag= ib
    /tag= ic
    /tag= id
    /tag= ie
    /tag= if
    /tag= ig
    /tag= ih
    /tag= ii
    /tag= ij
    /tag= ik
    /tag= il
    /tag= im
    /tag= in
    /tag= io
    /tag= ip
    /tag= iq
    /tag= ir
    /tag= is
    /tag= it
    /tag= iu
    /tag= iv
    /tag= iw
    /tag= ix
    /tag= iy
    /tag= iz
    /tag= ja
    /tag= jb
    /tag= jc
    /tag= jd
    /tag= je
    /tag= jf
    /tag= jg
    /tag= jh
    /tag= ji
    /tag= jj
    /tag= jk
    /tag= jl
    /tag= jm
    /tag= jn
    /tag= jo
    /tag= jp
    /tag= jq
    /tag= jr
    /tag= js
    /tag= jt
    /tag= ju
    /tag= jv
    /tag= jw
    /tag= jx
    /tag= jy
    /tag= jz
    /tag= ka
    /tag= kb
    /tag= kc
    /tag= kd
    /tag= ke
    /tag= kf
    /tag= kg
    /tag= kh
    /tag= ki
    /tag= kj
    /tag= kk
    /tag= kl
    /tag= km
    /tag= kn
    /tag= ko
    /tag= kp
    /tag= kq
    /tag= kr
    /tag= ks
    /tag= kt
    /tag= ku
    /tag= kv
    /tag= kw
    /tag= kx
    /tag= ky
    /tag= kz
    /tag= la
    /tag= lb
    /tag= lc
    /tag= ld
    /tag= le
    /tag= lf
    /tag= lg
    /tag= lh
    /tag= li
    /tag= lj
    /tag= lk
    /tag= ll
    /tag= lm
    /tag= ln
    /tag= lo
    /tag= lp
    /tag= lq
    /tag= lr
    /tag= ls
    /tag= lt
    /tag= lu
    /tag= lv
    /tag= lw
    /tag= lx
    /tag= ly
    /tag= lz
    /tag= ma
    /tag= mb
    /tag= mc
    /tag= md
    /tag= me
    /tag= mf
    /tag= mg
    /tag= mh
    /tag= mi
    /tag= mj
    /tag= mk
    /tag= ml
    /tag= mm
    /tag= mn
    /tag= mo
    /tag= mp
    /tag= mq
    /tag= mr
    /tag= ms
    /tag= mt
    /tag= mu
    /tag= mv
    /tag= mw
    /tag= mx
    /tag= my
    /tag= mz
    /tag= na
    /tag= nb
    /tag= nc
    /tag= nd
    /tag= ne
    /tag= nf
    /tag= ng
    /tag= nh
    /tag= ni
    /tag= nj
    /tag= nk
    /tag= nl
    /tag= nm
    /tag= nn
    /tag= no
    /tag= np
    /tag= nq
    /tag= nr
    /tag= ns
    /tag= nt
    /tag= nu
    /tag= nv
    /tag= nw
    /tag= nx
    /tag= ny
    /tag= nz
    /tag= oa
    /tag= ob
    /tag= oc
    /tag= od
    /tag= oe
    /tag= of
    /tag= og
    /tag= oh
    /tag= oi
    /tag= oj
    /tag= ok
    /tag= ol
    /tag= om
    /tag= on
    /tag= oo
    /tag= op
    /tag= oq
    /tag= or
    /tag= os
    /tag= ot
    /tag= ou
    /tag= ov
    /tag= ow
    /tag= ox
    /tag= oy
    /tag= oz
    /tag= pa
    /tag= pb
    /tag= pc
    /tag= pd
    /tag= pe
    /tag= pf
    /tag= pg
    /tag= ph
    /tag= pi
    /tag= pj
    /tag= pk
    /tag= pl
    /tag= pm
    /tag= pn
    /tag= po
    /tag= pp
    /tag= pq
    /tag= pr
    /tag= ps
    /tag= pt
    /tag= pu
    /tag= pv
    /tag= pw
    /tag= px
    /tag= py
    /tag= pz
    /tag= qa
    /tag= qb
    /tag= qc
    /tag= qd
    /tag= qe
    /tag= qf
    /tag= qg
    /tag= qh
    /tag= qi
    /tag= qj
    /tag= qk
    /tag= ql
    /tag= qm
    /tag= qn
    /tag= qo
    /tag= qp
    /tag= qq
    /tag= qr
    /tag= qs
    /tag= qt
    /tag= qu
    /tag= qv
    /tag= qw
    /tag= qx
    /tag= qy
    /tag= qz
    /tag= ra
    /tag= rb
    /tag= rc
    /tag= rd
    /tag= re
    /tag= rf
    /tag= rg
    /tag= rh
    /tag= ri
    /tag= rj
    /tag= rk
    /tag= rl
    /tag= rm
    /tag= rn
    /tag= ro
    /tag= rp
    /tag= rq
    /tag= rr
    /tag= rs
    /tag= rt
    /tag= ru
    /tag= rv
    /tag= rw
    /tag= rx
    /tag= ry
    /tag= rz
    /tag= sa
    /tag= sb
    /tag= sc
    /tag= sd
    /tag= se
    /tag= sf
    /tag= sg
    /tag= sh
    /tag= si
    /tag= sj
    /tag= sk
    /tag= sl
    /tag= sm
    /tag= sn
    /tag= so
    /tag= sp
    /tag= sq
    /tag= sr
    /tag= ss
    /tag= st
    /tag= su
    /tag= sv
    /tag= sw
    /tag= sx
    /tag= sy
    /tag= sz
    /tag= ta
    /tag= tb
    /tag= tc
    /tag= td
    /tag= te
    /tag= tf
    /tag= tg
    /tag= th
    /tag= ti
    /tag= tj
    /tag= tk
    /tag= tl
    /tag= tm
    /tag= tn
    /tag= to
    /tag= tp
    /tag= tq
    /tag= tr
    /tag= ts
    /tag= tt
    /tag= tu
    /tag= tv
    /tag= tw
    /tag= tx
    /tag= ty
    /tag= tz
    /tag= ua
    /tag= ub
    /tag= uc
    /tag= ud
    /tag= ue
    /tag= uf
    /tag= ug
    /tag= uh
    /tag= ui
    /tag= uj
    /tag= uk
    /tag= ul
    /tag= um
    /tag= un
    /tag= uo
    /tag= up
    /tag= uq
    /tag= ur
    /tag= us
    /tag= ut
    /tag= uu
    /tag= uv
    /tag= uw
    /tag= ux
    /tag= uy
    /tag= uz
    /tag= va
    /tag= vb
    /tag= vc
    /tag= vd
    /tag= ve
    /tag= vf
    /tag= vg
    /tag= vh
    /tag= vi
    /tag= vj
    /tag= vk
    /tag= vl
    /tag= vm
    /tag= vn
    /tag= vo
    /tag= vp
    /tag= vq
    /tag= vr
    /tag= vs
    /tag= vt
    /tag= vu
    /tag= vv
    /tag= vw
    /tag= vx
    /tag= vy
    /tag= vz
    /tag= wa
    /tag= wb
    /tag= wc
    /tag= wd
    /tag= we
    /tag= wf
    /tag= wg
    /tag= wh
    /tag= wi
    /tag= wj
    /tag= wk
    /tag= wl
    /tag= wm
    /tag= wn
    /tag= wo
    /tag= wp
    /tag= wq
    /tag= wr
    /tag= ws
    /tag= wt
    /tag= wu
    /tag= wv
    /tag= ww
    /tag= wx
    /tag= wy
    /tag= wz
    /tag= xa
    /tag= xb
    /tag= xc
    /tag= xd
    /tag= xe
    /tag= xf
    /tag= xg
    /tag= xh
    /tag= xi
    /tag= xj
    /tag= xk
    /tag= xl
    /tag= xm
    /tag= xn
    /tag= xo
    /tag= xp
    /tag= xq
    /tag= xr
    /tag= xs
    /tag= xt
    /tag= xu
    /tag= xv
    /tag= xw
    /tag= xx
    /tag= xy
    /tag= xz
    /tag= ya
    /tag= yb
    /tag= yc
    /tag= yd
    /tag= ye
    /tag= yf
    /tag= yg
    /tag= yh
    /tag= yi
    /tag= yj
    /tag= yk
    /tag= yl
    /tag= ym
    /tag= yn
    /tag= yo
    /tag= yp
    /tag= yq
    /tag= yr
    /tag= ys
    /tag= yt
    /tag= yu
    /tag= yv
    /tag= yw
    /tag= yx
    /tag= yy
    /tag= yz
    /tag= za
    /tag= zb
    /tag= zc
    /tag= zd
    /tag= ze
    /tag= zf
    /tag= zg
    /tag= zh
    /tag= zi
    /tag= zj
    /tag= zk
    /tag= zl
    /tag= zm
    /tag= zn
    /tag= zo
    /tag= zp
    /tag= zq
    /tag= zr
    /tag= zs
    /tag= zt
    /tag= zu
    /tag= zv
    /tag= zw
    /tag= zx
    /tag= zy
    /tag= zz
    /tag= aa
    /tag= ab
    /tag= ac
    /tag= ad
    /tag= ae
    /tag= af
    /tag= ag
    /tag= ah
    /tag= ai
    /tag= aj
    /tag= ak
    /tag= al
    /tag= am
    /tag= an
    /tag= ao
    /tag= ap
    /tag= aq
    /tag= ar
    /tag= as
    /tag= at
    /tag= au
    /tag= av
    /tag= aw
    /tag= ax
    /tag= ay
    /tag= az
    /tag= ba
    /tag= bb
    /tag= bc
    /tag= bd
    /tag= be
    /tag= bf
    /tag= bg
    /tag= bh
    /tag= bi
    /tag= bj
    /tag= bk
    /tag= bl
    /tag= bm
    /tag= bn
    /tag= bo
    /tag= bp
    /tag= bq
    /tag= br
    /tag= bs
    /tag= bt
    /tag= bu
    /tag= bv
    /tag= bw
    /tag= bx
    /tag= by
    /tag= bz
    /tag= ca
    /tag= cb
    /tag= cc
    /tag= cd
    /tag= ce
    /tag= cf
    /tag= cg
    /tag= ch
    /tag= ci
    /tag= cj
    /tag= ck
    /tag= cl
    /tag= cm
    /tag= cn
    /tag= co
    /tag= cp
    /tag= cq
    /tag= cr
    /tag= cs
    /tag= ct
    /tag= cu
    /tag= cv
    /tag= cw
    /tag= cx
    /tag= cy
    /tag= cz
    /tag= da
    /tag= db
    /tag= dc
    /tag= dd
    /tag= de
    /tag= df
    /tag= dg
    /tag= dh
    /tag= di
    /tag= dj
    /tag= dk
    /tag= dl
    /tag= dm
    /tag= dn
    /tag= do
    /tag= dp
    /tag= dq
    /tag= dr
    /tag= ds
    /tag= dt
    /tag= du
    /tag= dv
    /tag= dw
    /tag= dx
    /tag= dy
    /tag= dz
    /tag= ea
    /tag= eb
    /tag= ec
    /tag= ed
    /tag= ee
    /tag= ef
    /tag= eg
    /tag= eh
    /tag= ei
    /tag= ej
    /tag= ek
    /tag= el
    /tag= em
    /tag= en
    /tag= eo
    /tag= ep
    /tag= eq
    /tag= er
    /tag= es
    /tag= et
    /tag= eu
    /tag= ev
    /tag= ew
    /tag= ex
    /tag= ey
    /tag= ez
    /tag= fa
    /tag= fb
    /tag= fc
    /tag= fd
    /tag= fe
    /tag= ff
    /tag= fg
    /tag= fh
    /tag= fi
    /tag= fj
    /tag= fk
    /tag= fl
    /tag= fm
    /tag= fn
    /tag= fo
    /tag= fp
    /tag= fq
    /tag= fr
    /tag= fs
    /tag= ft
    /tag= fu
    /tag= fv
    /tag= fw
    /tag= fx
    /tag= fy
    /tag= fz
    /tag= ga
    /tag= gb
    /tag= gc
    /tag= gd
    /tag= ge
    /tag= gf
    /tag= gh
    /tag= gi
    /tag= gj
    /tag= gk
    /tag= gl
    /tag= gm
    /tag= gn
    /tag= go
    /tag= gp
    /tag= gq
    /tag= gr
    /tag= gs
    /tag= gt
    /tag= gu
    /tag= gv
    /tag= gw
    /tag= gx
    /tag= gy
    /tag= gz
    /tag= ha
    /tag= hb
    /tag= hc
    /tag= hd
    /tag= he
    /tag= hf
    /tag= hg
    /tag= hh
    /tag= hi
    /tag= hj
    /tag= hk
    /tag= hl
    /tag= hm
    /tag= hn
    /tag= ho
    /tag= hp
    /tag= hq
    /tag= hr
    /tag= hs
    /tag= ht
    /tag= hu
    /tag= hv
    /tag= hw
    /tag= hx
    /tag= hy
    /tag= hz
    /tag= ia
    /tag= ib
    /tag= ic
    /tag= id
    /tag= ie
    /tag= if
    /tag= ig
    /tag= ih
    /tag= ii
    /tag= ij
    /tag= ik
    /tag= il
    /tag= im
    /tag= in
    /tag= io
    /tag= ip
    /tag= iq
    /tag= ir
    /tag= is
    /tag= it
    /tag= iu
    /tag= iv
    /tag= iw
    /tag= ix
    /tag= iy
    /tag= iz
    /tag= ja
    /tag= jb
    /tag= jc
    /tag= jd
    /tag= je
    /tag= jf
    /tag= jg
    /tag= jh
    /tag= ji
    /tag= jj
    /tag= jk
    /tag= jl
    /tag= jm
    /tag= jn
    /tag= jo
    /tag= jp
    /tag= jq
    /tag= jr
    /tag= js
    /tag= jt
    /tag= ju
    /tag= jv
    /tag= jw
    /tag= jx
    /tag= jy
    /tag= jz
    /tag= ka
    /tag= kb
    /tag= kc
    /tag= kd
    /tag= ke
    /tag= kf
    /tag= kg
    /tag= kh
    /tag= ki
    /tag= kj
    /tag= kk
    /tag= kl
    /tag= km
    /tag= kn
    /tag= ko
    /tag= kp
    /tag= kq
    /tag= kr
    /tag= ks
    /tag= kt
    /tag= ku
    /tag= kv
    /tag= kw
    /tag= kx
    /tag= ky
    /tag= kz
    /tag= la
    /tag= lb
    /tag= lc
    /tag= ld
    /tag= le
    /tag= lf
    /tag= lg
    /tag= lh
    /tag= li
    /tag= lj
    /tag= lk
    /tag= ll
    /tag= lm
    /tag= ln
    /tag= lo
    /tag= lp
    /tag= lq
    /tag= lr
    /tag= ls
    /tag= lt
    /tag= lu
    /tag= lv
    /tag= lw
    /tag= lx
    /tag= ly
    /tag= lz
    /tag= ma
    /tag= mb
    /tag= mc
    /tag= md
    /tag= me
    /tag= mf
    /tag= mg
    /tag= mh
    /tag= mi
    /tag= mj
    /tag= mk
    /tag= ml
    /tag= mm
    /tag= mn
    /tag= mo
    /tag= mp
    /tag= mq
    /tag= mr
    /tag= ms
    /tag= mt
    /tag= mu
    /tag= mv
    /tag= mw
    /tag= mx
    /tag= my
    /tag= mz
    /tag= na
    /tag= nb
    /tag= nc
    /tag= nd
    /tag= ne
    /tag= nf
    /tag= ng
    /tag= nh
    /tag= ni
    /tag= nj
    /tag= nk
    /tag= nl
    /tag= nm
    /tag= nn
    /tag= no
    /tag= np
    /tag= nq
    /tag= nr
    /tag= ns
    /tag= nt
    /tag= nu
    /tag= nv
    /tag= nw
    /tag= nx
    /tag= ny
    /tag= nz
    /tag= oa
    /tag= ob
    /tag= oc
    /tag= od
    /tag= oe
    /tag= of
    /tag= og
    /tag= oh
    /tag= oi
    /tag= oj
    /tag= ok
    /tag= ol
    /tag= om
    /tag= on
    /tag= oo
    /tag= op
    /tag= oq
    /tag= or
    /tag= os
    /tag= ot
    /tag= ou
    /tag= ov
    /tag= ow
    /tag= ox
    /tag= oy
    /tag= oz
    /tag= pa
    /tag= pb
    /tag= pc
    /tag= pd
    /tag= pe
    /tag= pf
    /tag= pg
    /tag= ph
    /tag= pi
    /tag= pj
    /tag= pk
    /tag= pl
    /tag= pm
    /tag= pn
    /tag= po
    /tag= pp
    /tag= pq
    /tag= pr
    /tag= ps
    /tag= pt
    /tag= pu
    /tag= pv
    /tag= pw
    /tag= px
    /tag= py
    /tag= pz
    /tag= qa
    /tag= qb
    /tag= qc
    /tag= qd
    /tag= qe
    /tag= qf
    /tag= qg
    /tag= qh
    /tag= qi
    /tag= qj
    /tag= qk
    /tag= ql
    /tag= qm
    /tag= qn
    /tag= qo
    /tag= qp
    /tag= qq
    /tag= qr
    /tag= qs
    /tag= qt
    /tag= qu
    /tag= qv
    /tag= qw
    /tag= qx
    /tag= qy
    /tag= qz
    /tag= ra
    /tag= rb
    /tag= rc
    /tag= rd
    /tag= re
    /tag= rf
    /tag= rg
    /tag= rh
    /tag= ri
    /tag= rj
    /tag= rk
    /tag= rl
    /tag= rm
    /tag= rn
    /tag= ro
    /tag= rp
    /tag= rq
    /tag= rr
    /tag= rs
    /tag= rt
    /tag= ru
    /tag= rv
    /tag= rw
    /tag= rx
    /tag= ry
    /tag= rz
    /tag= sa
    /tag= sb
    /tag= sc
    /tag= sd
    /tag= se
    /tag= sf
    /tag= sg
    /tag= sh
    /tag= si
    /tag= sj
    /tag= sk
    /tag= sl
    /tag= sm
    /tag= sn
    /tag= so
    /tag= sp
    /tag= sq
    /tag= sr
    /tag= ss
    /tag= st
    /tag= su
    /tag= sv
    /tag= sw
    /tag= sx
    /tag= sy
    /tag= sz
    /tag= ta
    /tag= tb
    /tag= tc
    /tag= td
    /tag= te
    /tag= tf
    /tag= tg
    /tag= th
    /tag= ti
    /tag= tj
    /tag= tk
    /tag= tl
    /tag= tm
    /tag= tn
    /tag= to
    /tag= tp
    /tag= tq
    /tag= tr
    /tag= ts
    /tag= tt
    /tag= tu
    /tag= tv
    /tag= tw
    /tag= tx
    /tag= ty
    /tag= tz
    /tag= ua
    /tag= ub
    /tag= uc
    /tag= ud
    /tag= ue
    /tag= uf
    /tag= ug
    /tag= uh
    /tag= ui
    /tag= uj
    /tag= uk
    /tag= ul
    /tag= um
    /tag= un
    /tag= uo
    /tag= up
    /tag= uq
    /tag= ur
    /tag= us
    /tag= ut
    /tag= uu
    /tag= uv
    /tag= uw
    /tag= ux
    /tag= uy
    /tag= uz
    /tag= va
    /tag= vb
    /tag= vc
    /tag= vd
    /tag= ve
    /tag= vf
    /tag= vg
    /tag= vh
    /tag= vi
    /tag= vj
    /tag= vk
    /tag= vl
    /tag= vm
    /tag= vn
    /tag= vo
    /tag= vp
    /tag= vq
    /tag= vr
    /tag= vs
    /tag= vt
    /tag= vu
    /tag= vv
    /tag= vw
    /tag= vx
    /tag= vy
    /tag= vz
    /tag= wa
    /tag= wb
    /tag= wc
    /tag= wd
    /tag= we
    /tag= wf
    /tag= wg
    /tag= wh
    /tag= wi
    /tag= wj
    /tag= wk
    /tag= wl
    /tag= wm
    /tag= wn
    /tag= wo
    /tag= wp
    /tag= wq
    /tag= wr
    /tag= ws
    /tag= wt
    /tag= wu
    /tag= wv
    /tag= ww
    /tag= wx
    /tag= wy
    /tag= wz
    /tag= xa
    /tag= xb
    /tag= xc
    /tag= xd
    /tag= xe
    /tag= xf
    /tag= xg
    /tag= xh
    /tag= xi
    /tag= xj
    /tag= xk
    /tag= xl
    /tag= xm
    /tag= xn
    /tag= xo
    /tag= xp
    /tag= xq
    /tag= xr
    /tag= xs
    /tag= xt
    /tag= xu
    /tag= xv
    /tag= xw
    /tag= xx
    /tag= xy
    /tag= xz
    /tag= ya
    /tag= yb
    /tag= yc
    /tag= yd
    /tag= ye
    /tag= yf
    /tag= yg
    /tag= yh
    /tag= yi
    /tag= yj
    /tag= yk
    /tag= yl
    /tag= ym
    /tag= yn
    /tag= yo
    /tag= yp
    /tag= yq
    /tag= yr
    /tag= ys
    /tag= yt
    /tag= yu
    /tag= yv
    /tag= yw
    /tag= yx
    /tag= yy
    /tag= yz
    /tag= za
    /tag= zb
    /tag= zc
    /tag= zd
    /tag= ze
    /tag= zf
    /tag= zg
    /tag= zh
    /tag= zi
    /tag= zj
    /tag= zk
    /tag= zl
    /tag= zm
    /tag= zn
    /tag= zo
    /tag= zp
    /tag= zq
    /tag= zr
    /tag= zs
    /tag= zt
    /tag= zu
    /tag= zv
    /tag= zw
    /tag= zx
    /tag= zy
    /tag= zz

```

```

MO9010648-A.
PD 20-SEP-1990.
PF 14-MAR-1990; U01403.
PR 14-MAR-1989; US-321384.
PR 12-MAR-1990; US-492555.
PA (SALK ) SALK INST FOR BIOL STUD.
PI Heinemann SF; Deneris ES; Duvolsin RM; Patrick JW;
DR MPI; 90-304987/40.
P-PSDB; R07143.
PT New neuronal nicotinic acetylcholine receptor - compns. contg.
PT beta 4 sub-unit and DNA sequences encoding them.
PS Claim 11; Fig 3; 47pp; English.
CC The sequence encodes a novel neuronal nicotinic acetylcholine
CC receptor subunit, beta 4. Plasmid pPC13 (ATCC 67893) was
CC isolated from a cDNA library in lambda ZAP II prep. from rat
CC mRNA. The sequence of the cDNA in the region encoded by the 5th
CC exon is identical to that determined for the genomic clone with
CC the exception of a sustn. at posn. 720 of a t for a c (this does
CC not alter the AA sequence); this may be due to a polymorphism
CC between the rat strains used for the genomic library and the cDNA
CC library, or could have resulted from a reverse transcriptase error
CC during prep. of the cDNA. A sequence, the core of which is 44 bp
CC long is repeated three times at the beginning of the 3' UT region.
CC The function of this remains unknown. The protein subunit can
CC combine with the known subunits, alpha-2, -3, and -4, and beta-2
CC to form previously unknown functional receptors. The new subunit
CC is expressed in the central and peripheral nervous systems and in
CC PC12 cells.
CU Sequence 2460 BP; 519 A; 767 C; 587 G; 587 T;
SQ

```

```

Query Match 30.6%; Score 30.6; DB 1; Length 2460;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

OY 9 AGTGGTATGAGCGCTTGAGCGCCATTCACACTAAGCTGGTGGTAATCTCTCGG 68
DB 415 AATGCCAGTACCTGCTCCAGSTAAGCTGCAACATTTCACACAATGCTGGTGA 474
OY 69 CATGCCAGTACCTGCTCCAGSTAAGCTGCA 93
DB 475 AGCATCCAGTACCTGCTCCAGSTAAGCTGCA 499

```


RESULT 7
X23530 standard; DNA; 2192 BP.

AC X23530;
DE 17-JUN-1999 (first entry)
KW Maize Xa21 gene D74 DNA fragment.
KW Xa21: receptor kinase-like protein; multigene family; RRR; rice; D74;
KW plant disease resistance; cassava; maize; tomato; Xanthomonas; ss.
OS Zea mays.
PN WO9909151-A2.
PD 25-FEB-1999.
PF 17-JUL-1998; U14841.
PR 13-AUG-1997; US-910386.
PA (REGC) UNIV CALIFORNIA.
PI Hubert SH, Richter T, Ronald PC, Song W, Szabo V,
PI Wang G,
PI WPI: 99-204431/17.
PT New RRR polynucleotides and nucleic acid constructs - used for
PT generating transgenic plants resistant to Xanthomonas
PT Claim 12: Page 57-58; 67pp; English.
CC This invention describes a method for conferring disease resistance in
CC plants. The invention describes the use of novel genes and proteins
CC belonging to the Oryza longistaminata and Oryza sativa receptor
CC kinase-like protein (RRK) Xa21 multigene family. Such genes from
CC cassava, maize and tomato are also described. The genes and proteins can
CC be used for enhancing resistance to Xanthomonas in a plant, preferably in
CC rice or tomato.
SQ Sequence 2192 BP: 422 A; 659 C; 693 G; 418 T;

Query Match
Best Local Similarity 28.4%; Score 28.4; DB 1; Length 2192;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 13 CTGATGAGCGCTTTGAGCGGCACATTCACACTACGCTGTGGAATCTCTGGGCATT 72
DB 645 CTGCTGGCGCGGCTGCGCCAGCTCGCTGGAATGATGAGTGTGTCGACGTGACCTG 704
OY 73 GCCAGTACCTGCTCCCTCAGGTAACTG 98
DB 705 TCCAGCACCGACGTGGAAGGAACTG 730

RESULT 8
T79976/c
ID T79976 standard; cDNA; 1365 BP.
AC T79976;
DE 04-FEB-1998 (first entry)
KW Human galanin receptor GALT2 cDNA.
KW Galanin receptor; GALT2; human; agonist; antagonist; pain;
KW eating disorder; Alzheimer's disease; diabetes; ischemia;
KW obesity; bulimia; anorexia; nociception; cognitive disorder;
KW sensory disorder; motion sickness; hypertension; convulsion;
KW epilepsy; glaucoma; ulcer; inflammation; immune disease; anxiety;
KW therapy; ss.
OS Homo sapiens.
PN Key
FT CDS
FT Location/Qualifiers
FT 102..1265
FT /*tag- a
PD WO9726853-A2.
PD 31-JUL-1997.
PF 24-JAN-1997;
PF 27-SEP-1996; US-721837.
PR 24-JAN-1996; US-590494.
PR 01-APR-1996; US-626046.
PR 01-APR-1996; US-626885.
PA (SYNA-) SYNAPTIC PHARM CORP.
PI Branchek T, Forray C, Gerald CPG, Linemeyer D, Smith KE,
PI Weinshank R,
PI WPI: 97-393326/36.
PT P-PSDB: W24562.
PT Nucleic acid molecule encoding galanin receptor GALT2 - agonists and
PT antagonists of which are useful to treat eating disorders, pain and

PT Alzheimer's disease
PS Claim 56; Fig 10; 236pp; English.
CC This cDNA clone codes for a novel claimed galanin receptor subtype,
CC designated GALT2 (see W24562). It was isolated from a human heart,
CC cDNA library using primers based on rat GALT2 cDNA (see T79975).
CC Also claimed are: (1) an isolated nucleic acid molecule (1) encoding
CC a modified GALT2 having an amino acid deletion, replacement or
CC addition in the third intracellular domain; (2) a purified GALT2
CC protein; (3) a vector comprising (1); and (4) a cell comprising the
CC GALT2 by (claimed). GALT2 antagonists can be used to treat obesity,
CC bulimia or Alzheimer's disease, while GALT2 agonists can be used to
CC treat anorexia or pain, or to decrease nociception (claimed).
CC Agonists and antagonists can also be used to treat other disorders,
CC including cognitive disorders, sensory disorders, motion sickness,
CC hypertension, convulsion/epilepsy, diabetes, glaucoma, reproductive
CC disorders, gastric and intestinal ulcers, inflammation, immune
CC disorders and anxiety. Antibodies can be used to detect the
CC presence of GALT2 on the surface of a cell, while transgenic
CC mammals can be used to determine the physiological effects of
CC varying levels of GALT2 activity (claimed).
SQ Sequence 1365 BP: 184 A; 498 C; 424 G; 259 T;

Query Match
Best Local Similarity 26.6%; Score 26.6; DB 1; Length 1365;
Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 13 CTGATGAGCGCTTTGAGCGGCACATTCACACTACGCTGTGGAATCTCTGGGCATT 72
DB 1278 CCGCTAAGTCTTTCAGCGGCACATCAACCGTACGATGCTGCGCTTGGGCCCT 1219
OY 73 GCCAGTACCTGCG 85
DB 1218 GCCAGAGCGGCC 1206

RESULT 9
V10329/c
ID V10329 standard; cDNA; 1083 BP.
AC V10329;
DE 08-JUN-1998 (first entry)
KW Human GALT2 receptor partial cDNA.
KW Galanin receptor; GALT2; disease; treatment; screening; activity;
KW antinociceptive activity; obesity; stroke; smooth muscle contraction;
KW cardiovascular activity; diabetes; anorexia; pituitary hormone release;
KW Alzheimer's disease; neuropathic pain; endocrine disorder; antibody;
KW gene therapy; transgenic animal; human; ds.
OS Homo sapiens.
PN Key
FT CDS
FT Location/Qualifiers
FT 1..997
FT /*tag- a
FT /product- GALT2
FT /note- "partial coding region of galanin recepto"
PD WO9746681-A2.
PD 11-DEC-1997.
PF 05-JUN-1997; U09787.
PR 03-JUN-1997; US-665034.
PR 05-JUN-1996; US-665034.
PA (FAR) BAYER CORP.
PI Bloomquist BT, Cornfield LJ, Heeja Y, McCaleb ML,
PI WPI: 98-042189/04.
PT P-PSDB: W40137.
PT Galanin receptor protein GALT2 and encoding polynucleotide - useful
PT for e.g. identifying GALT2 receptor agonists or antagonists to
PT regulate physiological responses e.g. to treat obesity and diabetes
PS Claim 1; Fig 6; 44pp; English.
CC This cDNA sequence encodes a novel human galanin receptor protein, GALT2.
CC activity. Galanin has effects such as antinociceptive activity, smooth
CC muscle contraction, cardiovascular activity, pituitary hormone release,
CC cognition and increased food intake and identified drugs can regulate
CC physiological responses associated with the GALT2 receptor, e.g.

CC GALR2 gene was mapped to chromosome 17q25. The invention provides

cognitive dis

PT New mouse galanin receptor, GALR2, is useful to identify agonists and
PT antagonists to treat conditions involving galanin, e.g. for treating
PT obesity, pain or cognitive disorders
PS Example 6: Fig 7A-B: 55pp: English
CC This is the DNA sequence of the human GALR2 gene that codes for a
CC novel galanin receptor, designated GALR2 (see W61462), that is a
CC member of the G-protein coupled receptor family. It was isolated
CC from an EMBL3 Spe/NotI human genomic library using as probe a partial
CC gene sequence that had been amplified from genomic DNA by PCR using
CC primers (see V45092-93) based on transmembrane regions of
CC somatostatin and somatostatin-related gene receptors. The human
CC GALR2 gene was mapped to chromosome 17q25. The invention provides
CC rat, human and mouse GALR2 sequences (see W61461-63). Also
CC provided are nucleic acids encoding these novel receptor sequences
CC and assays to identify ligands particularly to murine GALR2. Such
CC ligands may be useful therapeutically e.g. to treat obesity or
CC cognitive disorders involving excess galanin or to treat pain or
CC anorexia involving insufficient galanin. Nucleic acids encoding

CC GALR2 are useful in assays for GALR2 and to produce hybridisation
CC probes to screen for similar receptors or for GALR2 in other species.
SQ Sequence 3390 BP; 571 A; 1132 C; 1025 G; 660 T;

Query Match
Best Local Similarity 26.6%; Score 26.6; DB 1; Length 3390;
Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 13 CTGATGAGCGCTTTGACGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 72
DB 2871 CCGCTAGTGTCTTCAGGCCACATCAACCGTCAGATGCTGTCCCTTCGGCCCT 2812

QY 73 GCCAGTACTGCC 85
DB 2811 GCCAGGACGGGCC 2799

RESULT 12
649/c
V32649 standard; DNA; 3390 BP.

~ V32649;
DT 24-NOV-1998 (first entry)
DE Human galanin receptor GALR2 genomic DNA.
KM Galanin receptor; GALR2 gene; human; ligand; obesity; anorexia;
KW pain; cognitive disorder; therapy; G-protein coupled receptor; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 305..2858
FT /tag- a
FT /note- "contains an intron"
FT /tag- b
FT /number- 1
FT /tag- c
FT /number- 1
FT /tag- d
FT /number- 11
PN 08-JUL-1998.
PD 18-DEC-1997; U23890.
PR 27-DEC-1996; US-033851.
PA (MERI) MERCK & CO INC.
PI Sullivan K, Tan C;
DR WPI: 98-388037/33.
P-PSDB: W49003.

New galanin receptor, GALR2 - useful, e.g. to identify agonists and
antagonists, therapeutically to treat conditions involving excess or
insufficient galanin such as obesity
PS Example 6; Fig 7A-B; 57pp; English.
CC This is the DNA sequence of the human GALR2 gene that codes for a
CC novel galanin receptor, designated GALR2 (see W49003), that is a
CC member of the G-protein coupled receptor family. It was isolated
CC from an EMBL3 Sf6/7 human genomic library using as probe a partial
CC gene sequence that had been amplified from genomic DNA by PCR using
CC primers (see V32652-53) based on transmembrane regions of
CC somatostatin and somatostatin-related gene receptors. The human
CC GALR2 gene was mapped to chromosome 17q25. The invention provides
CC rat, human and mouse GALR2 sequences (see W49002-04). Also
CC provided are nucleic acids encoding these novel receptor sequences
CC and assays to identify ligands particular to GALR2. Such ligands
CC may be useful therapeutically e.g. to treat obesity or anorexia
CC disorders involving excess galanin or to treat pain or anorexia
CC involving insufficient galanin. Nucleic acids encoding GALR2 are
CC useful in assays for GALR2 and to produce hybridisation probes to
CC screen for similar receptors or for GALR2 in other species.
SQ Sequence 3390 BP; 571 A; 1132 C; 1025 G; 660 T;

Query Match
Best Local Similarity 26.6%; Score 26.6; DB 1; Length 3390;
Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 13 CTGATGAGCGCTTTGACGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 72
DB 2871 CCGCTAGTGTCTTCAGGCCACATCAACCGTCAGATGCTGTCCCTTCGGCCCT 2812

QY 73 GCCAGTACTGCC 85
DB 2811 GCCAGGACGGGCC 2799

RESULT 13

ID X13051 standard; DNA; 13574 BP.
AC X13051;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:114.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046555.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 706-713; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13191 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 13574 BP; 4658 A; 2376 C; 2922 G; 3608 T;

Query Match
Best Local Similarity 25.2%; Score 25.2; DB 1; Length 13574;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 4 TTTCTAGTGTCTTCAGGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 63
DB 3962 TTTCAGTGTCTTCAGGCCATTCACACTACGTTGGTGAATTCCTTCGGCATT 4021

QY 64 CTGGCATTCG 73
DB 4022 TTGGTATCG 4031

RESULT 14

ID X25011 standard; DNA; 5002 BP.
AC X25011;
DT 05-JUL-1999 (first entry)
DE Anthr- specific Esj2a gene and promoter.
KW Esj2a gene; promoter; pollen; anther dehiscence; male sterile;
OS Arabidopsis thaliana.

FT	Key	Location/Qualifiers
FT	promoter	1..2178
FT		/*tag= a
FT	CDS	2179..4475
FT		/*tag= b
FT	exon	/note= "contains 8 introns"
FT		2179..2433
FT		/*tag= c
FT	intron	/number= 1
FT		2434..2622
FT		/*tag= d
FT	exon	/number= 1
FT		2623..2754
FT		/*tag= e
FT		/number= 2
FT	intron	2755..2838
FT		/*tag= f
FT		/number= 2
FT	exon	2839..3009
FT		/*tag= g
FT	intron	/number= 3
FT		3010..3154
FT		/*tag= h
FT	exon	/number= 3
FT		3155..3175
FT		/*tag= i
FT	intron	/number= 4
FT		3176..3287
FT		/*tag= j
FT	exon	/number= 4
FT		3288..3495
FT		/*tag= k
FT	intron	/number= 5
FT		3496..3629
FT		/*tag= l
FT	exon	/number= 5
FT		3630..3711
FT		/*tag= m
FT	intron	/number= 6
FT		3712..3873
FT		/*tag= n
FT	exon	/number= 6
FT		3874..3982
FT		/*tag= o
FT	intron	/number= 7
FT		3983..4072
FT		/*tag= p
FT	exon	/number= 7
FT		4073..4186
FT		/*tag= q
FT	intron	/number= 8
FT		4187..4271
FT		/*tag= r
FT	exon	/number= 8
FT		4272..4475
FT		/*tag= s
FT		/number= 9
FT		87..95
FT	primer_bind	/*tag= t
FT		/note= "primer pcr"
FT	primer_bind	770..792
FT		/*tag= u
FT	primer_bind	/note= "primer pcr"
FT		2164..2185
FT		/*tag= v
FT		/note= "primer 2A Nco"
PN	W09913089-A1.	
PD	18-MAR-1999.	
PF	11-SEP-1998; G02752.	
PR	11-SEP-1997; GB-019359.	
PA	(BIOG-) BIOGEMMA UK LTD.	
PI	Craze M, Paul W, Roberts JA;	
DR	WPI: 99-254279/21.	

DR P-50DB:W98178.
PT Generation of male sterile plants by controlling anther dehiscence
PS Example 1; Fig 3; 34pp; English.
CC This is the DNA sequence of the Esy2A gene of Arabidopsis thaliana.
CC The gene was isolated from an A. thaliana genomic library using
CC Brassica napus SMC66 cDNA as probe. The invention relates to the
CC use of the Esy2A promoter to reduce dehiscence and create male
CC sterile plants for use in hybrid seed production. The promoter is
CC used to drive expression of a further nucleic acid sequence that
CC results in prevention or reduction of anther dehiscence. For
CC example, expression of the RNase barnase causes cell ablation,
CC while expression of a plant hormone alters the developmental fate
CC of a cell. Plants are produced that have phenotypically normal
CC pollen grains, within phenotypically normal anthers, but in which
CC the anthers do not dehisce and thus do not release the pollen
CC grains. The system allows the female to be multiplied with the
CC artificial male sterility gene in the homozygous state, since the
CC female plant produces viable pollen. The system is suited to
CC crops which have high seed multiplication, large amounts of
CC pollen, and/or separate male and female inflorescences. These
CC factors allow for easy collection of pollen from non-dehiscing
CC anthers, facilitate self-pollination and the minimisation of the area
CC of plants that have to be self-pollinated manually. Such an
CC ideal crop is the monocot maize, but the system is also applicable
CC to e.g. wheat, barley, rice, fodder grass, banana, palm, orchid,
CC tulip, lily, melon, cucumber, tomato, pepper and willow. It is
CC also useful in the avoidance or reduction of pollen allergens and
CC may be effective in the control of asthma caused by pollen release.
SQ Sequence 5002 BP; 1702 A; 793 C; 821 G; 1686 T;

Query Match 24.8%; Score 24.8; DB 1; Length 5002;
Best Local Similarity 54.3%; Pred. No. 6.6;
Matches 50; Conservative 0; Mismatches 42; Indels 0; Gaps 0

OY 2 TGTTCCTAGTCTGATGACGGCTTTGACGCCACACTCCACACTAAAGTGTTGGTAGATTTC 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 TGTCGGCTAGTTGCTCGTAGAGACACTAAAGGTATGGTGCAGACTCATGTCCTTGAAGAATT 211
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 62 TTCCTGGCATTGCCCACTACCTGCCCTCCCAAGTA 93
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 210 ATCTAGTCATTCTTATTAACCTGGCCTTAGTTA 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
X16153
ID X16153 standard; cDNA to mRNA; 2376 BP.
AC X16153.
DT 22-APR-1999 (first entry)
DE Mouse Sox1 cDNA.
KW Mouse; Sox1; neuronal stem cell gene; neuroblastic cell; cancer;
KW nervous system; neurological disorder; ss.
NS Mus musculus.
OS W09900516-A2.
PD 07-JAN-1999.
PF 25-JUN-1998; G01862.
PR 25-JUN-1997; GB-013469.
PI (MEDTI-) MEDICAL RES COUNCIL.
FA Lovell-Badge R, Pevery LH, Smith A;
DR WPI: 99-095759/08.
PT Isolating neuroblastic cells from a population - by detecting the
PT expression of the sox1 gene in the cells and sorting the cells to
PT isolate those cells expressing Sox1
PS Disclosure, Page 43-44; 60pp; English.
CC A method has been developed for isolating neuroblastic cells from a cell
CC population. The method comprises: (a) detecting Sox1 gene expression in
CC the cells; and (b) isolating those cells expressing Sox1. Also described
CC is a method for producing a cell committed to the neuronal lineage,
CC comprising: (a) transfecting a pluripotent stem cell with a genetic
CC construct encoding Sox1 expression; and (b) culturing the stem cells to
CC differentiate into neural cells; and (c) isolating those neural cells
CC produced. The present sequence represents mouse Sox1 cDNA. Detection
CC of Sox1 expressing cells is important in diagnosing and treating cancers

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:23 ; Search time 2825.05 Seconds
(Without alignments)

69.823 Million cell updates/sec

SUMMARIES

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Title: US-08-956-518a-99
Perfect score: 100
Sequence: 1 CTGTTCTAGTGTCTGATGAG.....CTGCTCTCAGGTAAGCTGCA 100

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database:

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

Result No.	Score	Query Match	Length	DB	ID	Description
1	77.2	77.2	409	25	W03952	W03952 za62c08.r1
2	56.4	56.4	560	26	W52861	W52861 zc03e02.r1
3	38.6	38.6	607	44	AI292581	AI292581 GH1518.5
4	27.6	27.6	569	47	AI520119	AI520119 LD40102.3
5	27.4	27.4	689	48	AI547741	AI547741 PN3.1.01
6	27.2	27.2	318	28	AA083907	AA083907 zm62h12.r
7	27.2	27.2	322	28	AA112074	AA112074 zm62h12.s
8	27.2	27.2	440	28	AA123934	AA123934 mp78e03.r
9	27.2	27.2	139	30	AA267622	AA267622 mz98d02.r
10	27.2	27.2	408	33	AA422190	AA422190 v44e03.r
11	27.2	27.2	178	37	AA726126	AA726126 v08c07.r
12	26.4	26.4	458	31	AA726126	AA726126 v08c07.r
13	25.8	25.8	374	22	R72681	R72681 yj3d04.r1
14	25.8	25.8	216	29	AA138239	AA138239 mq78f03.r
15	25.8	25.8	522	46	AI450355	AI450355 mu28g03.x
16	25.6	25.6	350	26	W3186	W3186 zb35e08.r1
17	25.6	25.6	362	44	AD040432	AD040432 AD040432
18	25.6	25.6	274	49	AV016274	AV016274 AV016274
19	25.4	25.4	265	22	R35571	R35571 yb80g09.r1
20	25.4	25.4	456	22	R64451	R64451 y136g06.r1
21	25.4	25.4	484	49	AI635659	AI635659 tz08d02.x
22	25.2	25.2	255	22	R29672	R29672 FI-119AD.22
23	25.2	25.2	475	25	N53430	N53430 y225g02.s1
24	25.2	25.2	444	26	W95996	W95996 ze08b09.r1
25	25.2	25.2	197	29	AA184843	AA184843 mu49c05.r
26	25.2	25.2	416	32	AA355266	AA355266 EST63893
27	25.2	25.2	558	33	AA360428	AA360428 EST69822
28	25.2	25.2	464	40	AA398679	AA398679 zt70e02.s
29	25.2	25.2	407	41	AI004509	AI004509 O66h03.s
30	25.2	25.2	522	42	AI147987	AI147987 g963b07.s
31	25.2	25.2	406	44	AI312703	AI312703 qP83d02.x
32	25.2	25.2	188	49	AI627046	AI627046 mu49c05.y
33	25.2	25.2	500	35	AA590268	AA590268 vm19e02.r
34	25.0	25.0	614	43	AI177061	AI177061 EST720668
35	25.0	25.0	795	44	AI246505	AI246505 qm64a10.x
36	25.0	25.0	427	46	AI437303	AI437303 fd40a12.y
37	25.0	25.0	472	47	AI490165	AI490165 EST248796
38	25.0	25.0	471	47	AI490177	AI490177 EST248808
39	25.0	25.0	277	32	AA377412	AA377412 EST90023
40	24.8	24.8	277	42	AI15208	AI15208 UT-R-BTO-
41	24.8	24.8	623	47	AI487007	AI487007 EST245329
42	24.8	24.8	340	21	F11432	F11432 HSC2WH081.n
43	24.6	24.6	303	22	R46208	R46208 yj53e02.s1
44	24.6	24.6	411	24	N34208	N34208 yx77g03.r1
45	24.6	24.6				

ALIGNMENTS

RESULT 1
LOCUS W03952 409 bp mRNA
DEFINITION za62c08.r1 Soares fetal liver spleen INFILS Homo sapiens cDNA clone
IMAGE:297134.5' similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION W03952
NID g1273817

	sequence: 5'	CTCAGTTTTTTTTTTTTTTTTTTT	3'
BASE COUNT	93 a	67 c	80 g
ORIGIN		76 t	2 others

Query Match	27.28;	Score 27.2;	DB 28;	Length 318;
Best Local Similarity	64.18;	Pred. No. 6.3;		
Matches 41;	Conservative 0;	Mismatches 23;	Gaps 0	

QY 24 TTTSAGCGCACTTCCACACTAACGTTTGTAATCTTCGGGCATTGCCAGTACCTG 83
|||| | |||| | |||| | || ||| |||| | ||
Db 179 TTTGGCAGCACTTAAACACTCACGAATGAGGGCTATCCACGTCCCATGTACACGTCTC 128

Db 119 CTC 116

RESULT	7	AA112074	LOCUS	DEFINITION	ACCESSION
		AA112074	322 bp	mRNA	EST
			zmf6h12.s1	Stratigene fibroblast	23-DEC-1997
			IMAGE:530279 3'	mRNA sequence.	Hom sapiens cdna clone
			AA112074		

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS

1 (bases 1 to 322)
Hillier, J., Lennon, G., Becker, M., Bonaldi, M. T., Ch...
Eutheria; Primates; Catarrhini; Homidae; Homo.
Mammalia; Chordata; Vertebrata; Cnidaria; Mollusca; Arthropoda; Insecta; Annelida; Mollusca; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	On May 9, 1995 this sequence version replaced nt.002070

On May 8, 1995 this sequence version replaced g1:800954

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Email: est@watson.wustl.edu
 WABNTNC: <http://www.wabntnc.org>

plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

Full-length cDNAs available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28M13 rev2 from Amersham.

FEATURES
SOURCE

```

Location/Qualifiers
1. 318
/organism="Homo sapiens"
/db_xref="GDB:3919790"
/db_xref="taxon:9606"
/clone="IMAGE:530279"
/clone_1bp="Stratagene fibroblast (#937212)"
/lab_host="SOLR cells (Kanamycin resistant)"
/note="Vector: pUscript SK-7 Site 1: EcoRI. Site 2:
XhoI. Cloned unidirectionally. Primer: oligo dT. W138
cell line. Average insert size: 0.8 kb. UniZAP AR Vector
-5' adaptor sequence: 5' GAATTCGACACGAG 3'-3' adaptor

```

TITLE
JOURNAL
MEDLINE
COMMENT

On May 31, 1995 this sequence version replaced g1:802978

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: wilson@wustl.edu

WARNING: There is evidence

plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

FEATURES

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham.

Location/Qualifiers

FEATURES
SOURC

```

/organism="Homo sapiens"
/db_xref="GDB:3919790"
/db_xref="taxon:9606"
/map="1"
/clone="IMAGE:530279"
/clone_lib="Stratagene fibroblast (#937212)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Vector: pluescript SK+; site_1: EcoRI; site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. W138
cell line. Average insert size: 0.8 kb. Uni-ZAP XR Vector
-5' adaptor sequence: 5' GAAATCGACACAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTT 3'"
77 a 75 c 68 g 102 t

```


RESULT 14
 A138239 216 bp mRNA EST 09-FEB-1997
 LOCUS A138239
 DEFINITION mq78f03.f1 Stragagene mouse melanoma (#937312) Mus musculus cDNA
 clone IMAGE:584861 5', mRNA sequence.
 ACCESSION A138239
 VERSION 91700442
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 216)
 AUTHORS Marti,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMT Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced g1:1394413.

CONTACT: Maria M/Mouse EST Project
 WashU-HMT Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 66.
 Location/Qualifiers
 1..216
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="584861"
 /clone_lib="Stragagene mouse melanoma (#937312)"
 /tissue_type="melanoma"
 /dev_stage="M2 cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site: 1;
 EcorI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. From M2 cells, a highly metastatic derivative of
 the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCGACGAG
 3' -3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'."

BASE COUNT 83 a 39 g 39 c 55 t
 ORIGIN

Query Match 25.8%; Score 25.8; DB 29; Length 216;
 Best Local Similarity 58.4%; Pred. No. 17;
 Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTTGAGCGCATTCCACACTACGTTGGTGAATTCCTGGG 68
 115 AATGCTGATGTGTGTTATGCGCAGCTCCAACTACAGACTCAGTACATATTTCAA 174

69 CATTGCCAGTACCTGCC 85
 175 CAAGAAGAGTGTCAATC 191

RESULT 15
 A1450355 522 bp mRNA EST 09-MAR-1999
 LOCUS A1450355
 DEFINITION mu28g03.x1 Soares 2NBMt Mus musculus cDNA clone IMAGE:640756 3',

mRNA sequence.
 A1450355
 94296423
 VERSION A1450355.1 GI:4296423
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 522)
 AUTHORS Marti,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,
 Ritter,E., Korn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT On Jun 5, 1998 this sequence version replaced g1:3189570.

CONTACT: Maria M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 449.
 Location/Qualifiers
 1..522
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="640756"
 /clone_lib="Soares 2NBMt"
 /sex="male"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTAAGAGGAGCGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 174 a 106 c 103 g 139 t
 ORIGIN

Query Match 25.8%; Score 25.8; DB 46; Length 522;
 Best Local Similarity 58.4%; Pred. No. 24;
 Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

9 AGTGTGATGAGCGCTTTGAGCGCATTCCACACTACGTTGGTGAATTCCTGGG 68
 123 AATGCTGATGTGTGTTATGCGCAGCTCCAACTACAGACTCAGTACATATTTCAA 182

69 CATTGCCAGTACCTGCC 85
 183 CAAGAAGAGTGTCAATC 199

Search completed: September 17, 1999, 21:28:26
 Job time: 14290 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:30 ; Search time 2825.05 Seconds

(without alignments)
481.080 Million cell updates/sec

Title: US-08-956-518a-102

Sequence: 1 ACCCGCTTCCAGCGCGTAG.....CAAGAGTTCTCTACATCG 689

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database:

EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128.4	18.6	409	25	W03952
2	115.8	18.8	560	26	W52861
3	106	15.4	170	25	N83152
4	92.2	13.4	510	54	HS0011948
5	68.6	10.0	335	32	AA333244
6	49.4	7.2	607	44	A1292581
7	46.8	6.8	487	39	AA861176
8	42.2	6.1	274	34	AA460018
9	42.2	6.1	436	39	AA844642
10	42.2	6.1	396	45	A1367541
11	42.2	6.1	501	42	A1655990
12	41.8	6.1	521	53	HS0003687
13	40.8	5.9	494	37	AA697326
14	39.4	5.7	439	47	A1492967
15	39	5.7	557	43	A1192683
16	38.6	5.6	369	29	AA156428
17	37.8	5.5	354	26	W86837
18	37.8	5.5	790	48	A1585069
19	37	5.4	473	33	AA407186
20	37	5.4	446	44	A1317059
21	36.6	5.3	319	33	AA408177
22	36.4	5.3	400	50	A1681298
23	36.4	5.3	399	51	AU067986
24	36	5.2	435	42	A1085561
25	36	5.2	654	45	A1346522
26	36	5.2	373	47	A1462417
27	36	5.2	373	47	A1462417
28	36	5.2	415	24	H81299
29	35.6	5.2	289	37	AA702650
30	35.6	5.2	391	48	A1586795
31	35.6	5.2	605	40	AA979965
32	35.4	5.1	317	46	A1443300
33	35.4	5.1	638	26	W27222
34	35.2	5.1	232	20	D22064
35	35	5.1	232	20	D22108
36	35	5.1	216	20	D24865
37	35	5.1	443	27	AA040321
38	35	5.1	482	39	AA894793
39	35	5.1	415	45	A1363984
40	35	5.1	381	46	AA824494
41	35	5.1	482	49	A1660583
42	34.8	5.1	528	49	A1621763
43	34.8	5.1	630	49	A1622084
44	34.8	5.1	571	49	A1622684
45	34.8	5.1			

ALIGNMENTS

RESULT 1
LOCUS W03952
DEFINITION za62c08.r1 Soares fetal liver spleen INFUS Homo sapiens CDNA clone
IMAGE:297134 5' similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION W03952
NID 91275817

VERSION	W03952.1	GI:1275817
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Primates; Catarrhini; Homidae; Homo.	
	1 (bases 1 to 409)	
	Hillier,L., Clark,N., Dubucq,T., Elliston,R., Hawkins,M.,	
	Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,	
	Pearson,J., Ritzin,L., Rohlfing,T., Soares,M., Tan,F.,	
	Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and	
	Wilson.R.	
TITLE	The Mashu-Merck EST Project	
JOURNAL	Unpublished (1995)	
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LBNL; contact the Seq Consortium (info@image.llnl.gov) for further information. Seq primer: mob.REGA+R High quality sequence stop: 263. Location/Qualifiers 1..409 /organism="Homo sapiens" /db_xref="GDB:1242055" /db_xref="taxon:9606" /clone IMAGE:297134" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGAAATTAATTAAACATCTTTTTTTTTTTTTT 3']/ double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bernaldo."	
BASE COUNT	95 a 80 c 75 g 145 t 14 others	
OP'TIN		
Query Match	18.6%; Score 128.4; DB 25; Length 409;	
Best Local Similarity	92.9%; Pred. No. 1e-21; Indels 1; Gaps 1;	
Matches 143; Conservative	0; Mismatches 10; Indels 1; Gaps 1;	
OY 534	AATTCGAATTCGTATCA-GCATTTGTGGATGCTGCACAACCTGCATATTCGTATGAG 592	
DB 157	AGTCCATTCCTCTATCAGCAGCATTTNATGCTGCACAACCTGCATATTCGTATGAG 216	
OY 593	CGCTTTGACGCCACATTCACACATAACGTTGTGTAATCTTGCGCATTCGCCAGTAC 652	
DB 217	CGCTTTGACGCCACATTCACACATAACGTTGTGTAATCTTGCGCATTCGCCAGTAC 276	
OY 653	CTGCTCCAGCGCATATTCAGAAGTTCCTGCTGACA 666	
DB 277	CTGCTCCAGCGCATATTCAGAAGTTCCTGCTGACA 310	
RESULT 2		
LOCUS	W52861	
DEFINITION	W52861 560 bp mRNA EST 10-OCT-1996	
IMAGE:	zco3e02.r1 Soares-parathyroid.tumor.NBHP.Homo sapiens cDNA clone	
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN); mRNA sequence.	IMAGE:z2142.5 similar to gb:XY0297 NEURONAL ACTIVITYCHOLINE	
ACCESION	W52861	
NTD	g1350351	

	KEYWORDS	EST.	w52861.1	gi:1350351
	SOURCE	human.		
	ORGANISM	Homo sapiens		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	AUTHORS	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
		1 (bases 1 to 560)		
		Hillier L., Clark N., Dubucque T., Elliston R., Hawkins M.,		
		Hollman M., Hultman M., Kuchaba T., Le M., Lennon G., Marra M.,		
		Pearson V., Rifkin L., Rohlfing T., Soares M., Tan F.,		
		Trevaskis E., Waterston R., Williamson A., Woldmann P. and		
		Wilson R.		
	TITLE	The Mashu-Merck EST Project		
	JOURNAL	Unpublished (1995)		
	COMMENT	On Jan 24, 1995 this sequence version replaced gi:634363.		
		Contact: Wilson RK		
		Washington University School of Medicine		
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
		Tel: 314 286 1800		
		Fax: 314 286 1810		
		Email: est@wustl.wustl.edu		
		This clone is available royalty-free through LINDA; contact the		
		IMAGE Consortium (info@image.llnl.gov) for further information.		
		Insert Length: 1753 Std Error: 0.00		
		Seq primer: mob.REGA+ER		
		High quality sequence stop: 408.		
	FEATURES	Location/Qualifiers		
	source	1..560		
		/organism="Homo sapiens"		
		/db_xref="GDB:1258900"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:321242"		
		/clone_lib="Soares parathyroid_tumor_NbHPA"		
		/tissue_type="parathyroid tumor"		
		/dev_stage="adult"		
		/lab_host="DH10B (ampicillin resistant)"		
		/note="Organ: Parathyroid gland; Vector: pRTT3D		
		(Pharmacia) with a modified polylinker; Site_1: Not I;		
		Site_2: Eco RI; 1st strand cDNA was primed with a Not I -		
		[5'-		
		TTTTCACCATCTGAAGTGCGAGGCCGCACACATTTTTTTTTTTTTTTTTT		
		T-3'), double-stranded cDNA was size selected, ligated to		
		Eco RI adapters (Pharmacia), digested with Not I and		
		cloned into the Not I and Eco RI sites of a modified pRTT3D		
		vector (Pharmacia). Library went through one round of		
		normalization to a Cot = 5. Library constructed by Bento		
		Soares and M. Palma Bonaldo. RNA from sporadic parathyroid		
		adenomas was kindly provided by Dr. Stephen Marx, National		
		Institute of Diabetes and Digestive and Kidney Diseases,		
		NIH."		
	BASE COUNT	128 a 137 c 137 g 152 t 6 others		
	ORIGIN			
	Query Match	16.8%; Score 115.8; DB 26; Length 560;		
	Best Local Similarity	95.5%; Pred. No. 1,3e-18;		
	Matches 150; Conservative	0; Mismatches 4; Indels 3; Gaps 3		
QY	533	CATTCCATTGCTAAATCAGACTTTGTGGATAGCTCAAGAATGCATATTCGTGATGAG	592	
Db	37	CAGTTCACATTCGCAATTCACACATTTGGTAGTACTCCAACGTGATATTCGTGATGAG	96	
QY	593	GCGTTGAGGCGCACATTCACACACTAACGTGTGGTGAATTCCTCGGGCAATTCGCAGTAC	652	
Db	97	GCGTTGAGC-KNCAATTCACACACTAACGTGTGGTGAATTCCTCGGGCAATTCG-CAGTAC	154	
QY	653	CTGCCTCCAGGAGCATATTCAGAAGTTCCTGCTACATCG	689	
Db	155	CTGCCTCCCA-GNAATATTCAGAAGTTCCTGCTACATCG	190	


```

FEATURES      1. .170
source        Location/Qualifiers
```

Query Match	15.48;	Score 106;	DB 25;	Length 170;
Best Local Similarity	82.98;	Pred. No. 2.1e-16;		
Matches 121; Conservative	0.	Mismatch		

RESULT	4
HSM011948	
ID	HSM011948
XX	standard; RNA; EST; 510 BP.

```

XX Homo sapiens mRNA; EST DKFZps86b2218_r1 (from clone
DE DKFZps86b2218)
XX
XX
XX EST: expressed sequence tag.
XX
XX Homo sapiens (human)
OS
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria,
CC Primates; Catarrhini; Homnidae; Homo.
XX
XX [1]
XX 1-510
XX
XX Koehrer K., Beyer A., Mewes W., Gassenhuber J., Wiemann S.;
RA
RT
RL Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
CC
CC MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
XX
CC Clone from S. Wiemann, sequenced by BMF within the CDNA
CC sequencing consortium of the German genome Project
CC No s1 sequence available
CC
CC This clone is available at the RZPD in Berlin
CC
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
XX Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX
XX Key
XX
XX Location/Qualifiers
XX
XX source
XX 1..510
XX

```

Query Match	13.48;	Score 92.2;	DB 54;	Length 510;
Best Local Similarity	73.78;	Pred. No. 6.5e-13;		
Matches 154; Conservative	0;	Mismatches 33		

RESULT	5
AA333244	
LOCUS	AA333244
DEFINITION	EST737317 Embryo 8 week + "
	335 bp mRNA
	EST
	21-APR-199

ACCESSION	Similar to Interleukin-2 receptor, alpha chain, kappa B binding protein, mRNA sequence.
NID	AA333244
VERSION	91985508
KEYWORDS	AA333244.1 GI:1985508
SOURCE	EST
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
CDNA Library Arrayed by: Greg Lennon, Ph.D.


```

Db      256 TTCCATTGCTATCCAGCATTTGCGGATAGTCCAAAGTGGATAT 210
RESULT 12
ID      HSM003697 standard; RNA; EST; 521 BP.
XX      AL039221;
XX      AL039221.1
XX      e1397556
XX      e1397556
NI      12-MAR-1999 (Rel. 59, Created)
DT      12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX      Homo sapiens mRNA; EST: DKFZP(null),E131_f1 (from clone
DE      DKFZP(null),E131)
EST: expressed sequence tag.
..4      Homo sapiens (human)
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC      Primates; Catarrhini; Homnidae; Homo.
XX      [1]
XX      1-521
XX      Poustka A., Klein M., Mewes W., Gassenhuber J., Wiemann S.;
RA      Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RT      MITS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
RL      CLONE from S. Wiemann, sequenced by DKFZ within the CDNA
XX      sequencing consortium of the German Genome Project
CC      No s1 sequence available
CC      This clone is available at the RZPD in Berlin
CC      Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059
CC      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1..521
FH      /db_xref="taxon:9606"
FT      /organism="Homo sapiens"
FT      /clone="DKFZP(null),E131"
FT      /clone_11b="HU Drosophila melanogaster head Bluescript"
FT      DH10B; sites NotI + SalI
FT      /dev_stage="adult"
FT      /tissue_type="breast cancer"
..f      Sequence 521 BP; 62 A; 193 C; 176 G; 88 T; 2 other;
XX      S0
Query Match      6.1%; Score 41.8; DB 53; Length 521;
Best Local Similarity 47.4%; Pred. No. 1.1;
Matches 156; Conservative 0; Mismatches 169; Indels 4; Gaps 1;
QY      53 GCGGCTTACGAGGAGCCGCGCTCTGCTGAGTGGATTAAGAAGAGAGAGCCCG 112
Db      157 GAGGGGTGACAGGCGCGGGTCCCTTCACAGTGCTCTGCTTCAGGGCAAGCTCC 216
QY      113 CAGGCGCCACATGACTCCCGCCCAAGTCCCTCGGTGCGCCCTTTCAGAGCGCGCT 172
Db      217 GGTCTCCGGGCGCACTTNCCTGCTGCTGTCGNCATCTCTTCTCCAGCCTCTC 276
QY      173 CCGACGAGGAGTCAAGGCGCGGGAGAGTGAAGCGGCAAGCTCGGCGGGGGCCCG 232
Db      277 CCTTCGACAGGTGGATCTCGGTGGAGCGGAGCGGGGCGGGCGCC---CCGG 332
QY      223 CCGTGTGCGCGCGGCGGATGACAGAGGCTCGGAGACTGCTCTTTCCGCGCCCTCCCG 292
Db      333 GACCATGGCGGGGTCGACACCGCGCCCTCTCTAGCAGGCGGATGACCGCGAGCAG 392

```

```

QY      293 CCGAGGTGAGGGAAGATGTCATGTCAAGGTTCAAGGCCAAACCGAAGTACTGCGCTC 352
Db      393 GCCAGTGCCTGGCCACCCCGGGGTTCACAGGCTCCAGGGGAACCGAAGTCCGAGAGCC 452
QY      353 TATCTTCAGAGAGAACCGAGGCCACAGC 381
Db      453 CGAGGTCCCGGACCGAGAGGGGCTGCAGC 481
RESULT 13
AA697326 494 bp mRNA EST 28-NOV-1998
LOCUS      AA697326
DEFINITION      H102253.Sprine HL Drosophila melanogaster head Bluescript
Drosophila melanogaster cDNA clone H102253 Sprime similar to
M20316: D.melanogaster acetylcholine receptor-related protein mRNA,
complete cds, mRNA sequence.
ACCESSION      AA697326
VERSION      92700255
NID      AA697326.1 GI:2700255
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 494)
AUTHORS      Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
BROCKSTEIN, P., Lewis, S. and Rubin, G.M.
BDFP/HMT Drosophila EST Project
TITLE      Unpublished (1997)
JOURNAL      On Nov 20, 1997 this sequence version replaced gi:1172316.
COMMENT
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 22 row: E column: 5
High quality sequence stop: 416.
FEATURES
source      1..494
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="H102253"
/clone_11b="HL Drosophila melanogaster head Bluescript"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/note="Organ: head--brain & sensory organ; Vector:
Bluescript SK; Site: 1; EORI: Site 2; XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis Kit. Oligo dT-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK(+/-)"
BASE COUNT      124 a 114 c 127 g 129 t
ORIGIN
Query Match      5.9%; Score 40.8; DB 37; Length 494;
Best Local Similarity 59.5%; Pred. No. 1.9;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY      574 CTCGATATGCTGATGAGGCTTGACGCCACATTCACACTGATGTTGGAATTC 633
Db      244 CTTAATTAATGCGATGCACTACGAGTGGCTACAAATCCAACTGCTGATTATTC 303
QY      634 TTCTGGCATTCGCACTGCTCCAGGCAATTAAGAGTTCTCTGATACATCG 689
Db      304 CACGGAGAGAGTCTCTGTTGCTCTCGGCAATTACCAAGCTCTGCAACATCG 359
RESULT 14

```


A1492967
 LOCUS 439 bp mRNA EST 18-MAR-1999
 DEFINITION g247c02.x1 NCI-CGAP_Kid11 Homo sapiens CDNA clone IMAGE:2030018 3' similar to SW:CDNC_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR 1C; contains TARI.12 TARI repetitive element ;, mRNA sequence.
 ACCESSION A1492967
 NID 94393970
 VERSION A1492967.1 GI:4393970
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1. (bases 1 to 439)
 NCICGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 18, 1998 this sequence version replaced gi:3136536.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/db/rp/image/image.html

Insert Length: 908 Std Error: 0.00
 Seq primer: 40UP from Glbco
 High quality sequence stop: 65.
 Location/Qualifiers

FEATURES
 source 1. 439

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="21922.1-3"
 /clone="IMAGE:2030018"
 /clone_lib="NCI-CGAP_Kid11"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 132376-132391, 145607-145675, and 150952-150985). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 60 a 145 c 187 g 47 t
 ORIGIN

Query Match 5.7%; Score 39.4; DB 47; Length 439;
 Best Local Similarity 53.3%; Pred. No. 4;
 Matches 105; Conservative 0; Mismatches 91; Indels 1; Gaps 1;

QY 138 GTCCTGCGTCCCTTCCATTTCCAGCGCGCTCCACGAGGGTACGCGCGCGGGA 197
 Db 46 GCGCGCGCTGCGCTGTTGCGCGCGCTGCGCGCTGTTGAGGCGCGCGCGG 105
 QY 198 GAGGTGAGCGCGAGAGCTCGCGCGCGCGCGCGCTGTTGCGCGCGCAAGACGG 257
 Db 106 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT-GCTGGGCGCGCGCGCG 164
 QY 258 GCTGGGAGTGGCTCTTTTCCGCGCGCGCGCGCGAGAGTGAAGATGTCATG 317
 Db 165 ACTGGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGACCGAG 224

QY 318 TCAGGTTCAAGCCCA 334
 Db 225 ACCGCGACCGAGCGCA 241
 RESULT 15
 A1192683
 LOCUS 557 bp mRNA EST 29-OCT-1998
 DEFINITION g62607.x1 Soares_fetal_lung_NbHL19W Homo sapiens CDNA clone IMAGE:1743565 3' similar to SW:CDNC_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR 1C; contains MER22.b3 TARI repetitive element ;, mRNA sequence.
 ACCESSION A1192683
 NID 93743892
 VERSION A1192683.1 GI:3743892
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1. (bases 1 to 557)
 NCICGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1797633.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 862 Std Error: 0.00
 Seq primer: 40UP from Glbco
 High quality sequence stop: 118.
 Location/Qualifiers

FEATURES
 source 1. 557

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1743565"
 /clone_lib="Soares_fetal_lung_NbHL19W"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: lung; Vector: p773D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAGAGCGAGCGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHL19W."

BASE COUNT 62 a 193 c 242 g 60 t
 ORIGIN

Query Match 5.7%; Score 39; DB 43; Length 557;
 Best Local Similarity 53.3%; Pred. No. 5.4;
 Matches 104; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

QY 138 GTCCTGCGTCCCTTCCATTTCCAGCGCGCTCCACGAGGGTACGCGCGCGGGA 197
 Db 68 GCGCGCGCTGCGCTGTTGCGCGCGCTGCGCGCTGTTGAGGCGCGCGCGG 127
 QY 198 GAGGTGAGCGCGAGAGCTCGCGCGCGCGCGCGCTGTTGCGCGCGCAAGACGG 257
 Db 128 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG-6GCTGGGCGCGCGCG 186
 QY 258 GCTGGGAGTGGCTCTTTTCCGCGCGCGCGCGAGAGTGAAGATGTCATG 317

Db 187 ACTGAGCCGGGCGCGAGCCGAGCCGCGGGGCGGGGCGGAGGAGCCCG 246
OY 318 TCAGGGTTCAAGCC 332
Db 247 ACCGCGAGCCGAGCC 261

Search completed: September 17, 1999, 21:28:33
Job time: 14297 sec

Run on: September 17, 1999, 22:04:21 ; Search time 1962.17 seconds

29.9/0 Million cell updates/sec

Title: US-08-956-518A-100
Perfect score: 37

Sequence: 1 ACCCACACAGGCATATTCAGAGTTCCTGCTACATCG 37

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database

GeneEmbl.*
1: gb_bal.*
2: gb_bal2.*
3: gb_com.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_v1.*
18: em_fun.*
19: em_hcg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_com.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_v1.*
34: gb_hcg1.*
35: gb_hcg2.*
36: gb_in1.*
37: gb_in2.*
38: em_bal.*
39: em_bal2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	No.	Score	Query Match	Length	DB	ID	Description
1	30	81.1	1559	9	HUMA7NNR		L25827 Human a7 n

2	30	81.1	2087	10	HBARA7A
3	30	81.1	1599	10	HSNACHRA7
4	30	81.1	1977	10	HSU40583
5	30	81.1	1876	10	HSUG62436
6	30	81.1	641	11	AFO29838
7	30	81.1	689	11	AFO29839
8	30	81.1	1712	11	AFO36903
9	30	81.1	1836	11	AFO37646
10	28.6	77.3	1848	12	MOSNRAS
11	28.4	76.8	1855	12	MOSNRAS
12	73.0	2106	12	RATNARAD	
13	27	73.0	3030	12	S53987
14	25.4	68.6	1851	13	BRTX7
15	22.2	60.0	2251	10	HSY13931
16	22.2	60.0	191111	11	AC006080
17	22.2	60.0	171865	35	AC007284
18	22	59.5	2696	4	GGA7NAREC
19	22	59.5	2037	4	GGNNARA7A
20	21.8	58.9	335	11	HSTECTA16
21	21.6	58.4	194525	10	HS323M22
22	21.6	58.4	208613	11	ASC05539
23	21.6	58.4	276261	34	HSAC000406
24	21.2	57.3	143068	10	HSUY9567
25	21.2	57.3	44062	11	AC004517
26	21.2	57.3	1046	12	RATCHENS
27	21.2	57.3	1931	12	RATCNAR
28	21	56.8	1936	8	BTACHRECA
29	21	56.8	2082	9	HUMNACHR44
30	21	56.8	1597	10	HSCHRNA45
31	21	56.8	1884	10	HSNACHRA4
32	21	56.8	1833	10	HSNACHRA4
33	21	56.8	3343	10	HSUG62433
34	21	56.8	2149	12	AFO07212
35	21	56.8	2327	12	RATNARAB
36	21	56.8	2095	12	RATNARAC
37	20.8	56.2	51953	8	LEUB1378
38	20.8	56.2	87869	9	HS268B10
39	20.6	55.7	79976	7	ABO08267
40	20.6	55.7	99411	11	AC005207
41	20.6	55.7	312403	35	AC007541
42	20.6	55.7	3629	37	AFI43846
43	20.2	54.6	155379	10	HS61688
44	20.2	54.6	147419	10	HS738P11
45	20.2	54.6	177142	11	AC004582
					X70297 H.sapiens p
					Y08420 H.sapiens m
					Y040563 Human alpha
					U62426 Human nicot
					AFO29838 Homo sapl
					AFO29839 Homo sapl
					AFO36903 Homo sapl
					AF037646 Homo sapl
					L37663 Mus musculu
					Z23141 H.sapiens C
					L31619 Rattus ratt
					S53987 nicotinic r
					X93604 B.taurus mr
					Y13931 Homo saplen
					AC006080 Homo sapl
					AC007284 Homo sapl
					X52295 Chicken alp
					X68566 G.gallus mr
					AF055129 Homo sapi
					AL022476 Human DNA
					AC005539 Homo sapi
					AC000406 Homo sapi
					U95626 Homo saplen
					AC004517 Homo sapi
					M02096 Rat neurona
					L10077 Rat neurona
					X57032 B.taurus mr
					L35901 Human nicot
					X88745 H.sapiens m
					Y08421 H.sapiens m
					X87629 H.sapiens m
					U62433 Human nicot
					AFO07212 Rattus no
					M15683 Rat nicotin
					L31620 Rattus novr
					U81319 Lycopersico
					Z82244 Homo saplen
					ABO08264 Arabidops
					AC005207 Homo sapl
					AF007541 Homo sapl
					AFI43846 Heliothis
					AL02380

ALIGNMENTS

RESULT	1
HMA7NAR	
LOCUS	
DEFINITION	Human a7 nicotinic acetylcholine receptor mRNA.
ACCESSION	L28827
NID	9438616
VERSION	125827.1
KEYWORDS	GI:438616
SOURCE	nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.
ORGANISM	Human sapiens (library: Clontech HLU065b; ATCC 37433) fetus brain cDNA to mRNA.
REFERENCE	Homo sapiens (library: Clontech HLU065b; ATCC 37433) fetus brain cDNA to mRNA.
AUTHORS	Homo sapiens
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; 1 (sites)
JOURNAL	Dovecette-Stamm, L., Monteggia, L.M., Donnelly-Roberts, D., Wang, M.T.
FEATURES	Lee, J., Tian, J. and Giordano, T.
source	Cloning and sequence of the human a7 nicotinic acetylcholine receptor
	Drug Dev. Res. (1993) in press
	location/Qualifiers
	1..1559
	/organism="Homo sapiens"
	/db_xref="taxon:9606"


```

/dev_stage="fetus"
/tissue_type="brain"
/tissue_lib="Clontech HLI065b; ATCC 37433"
BASE COUNT      314 a      463 c      440 g      342 t
ORIGIN

Query Match      81.1%; Score 30; DB 9; Length 159;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CAGCATATTCAGAGTTCCTGCTACATCG 37
        |||||||
Db      405 CAGCATATTCAGAGTTCCTGCTACATCG 434

RESULT 2
        A7A      HSBARA7A      2087 bp      mRNA      PRI      01-JUN-1994
DEFINITION      H.sapiens mRNA for neuronal nicotinic acetylcholine receptor
                alpha-7 subunit.
ACCESSION      X70297
VERSION      X70297.1 GI:496606
KEYWORDS      neuronal nicotinic acetylcholine receptor alpha-7 subunit.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 2087)
AUTHORS      Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.
TITLE      Human alpha 7 acetylcholine receptor: cloning of the alpha 7
                subunit from the SH-SY5y cell line and determination of
                pharmacological properties of native receptors and functional alpha
                7 homomers expressed in Xenopus oocytes
JOURNAL      Mol. Pharmacol. 45 (3), 546-554 (1994)
MEDLINE      94195283
REFERENCE      2 (bases 1 to 2087)
AUTHORS      Katz, M.
TITLE      Direct Submission
JOURNAL      Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of
                Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235
                Stemmler Hall, Philadelphia, PA 19104, USA
FEATURES
        source
        1..2087
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /tissue_type="brain"
        /cell_type="neuroblastoma"
        /cell_line="SH-SY5y"
        /clone_lib="lambda zap II"
        /clone="SHS3"
        104..175
        /product="neuronal nicotinic acetylcholine receptor
                alpha-7 subunit"
        /codon_start=1
        /product="neuronal nicotinic acetylcholine receptor
                alpha-7 subunit"
        104..1612
        /product="neuronal nicotinic acetylcholine receptor
                alpha-7 subunit"
        /protein_id="CA449778.1"
        /db_xref="PID:9496607"
        /db_xref="GI:496607"
        /db_xref="SWISS-PROT:P36544"
        /translation="MRCSPGGVWGLAASLLHVSLOGEFOKRLKELYNPNLEPVP
                ANDSPLVIFSLIQIMDYDEKNQVLTNTIMWLSQMTDHYLQMNVSSEYGVKTVRF
                PDQIWKPDILLNSADERFDRFTNNVNVNSGHCQYLPGLFKSSCYIDVRFEPD
                VOCHKIKGWSYSGWSLDQMOEADISGITPNGBMDLVGIPGRKSERFECCKEYFP
                DVETFTVRRRTIYGLNLIPCVLISALALVFLPADSGEKISGITVLLSLVEM
                LVAETMPATSDSVPLAIFYASTNIIVGLSVVTVIVLQYHHDPDGCKMPKWTWVI
                LLMNCAMFLRMKRPEDKVRPAQCKORCSLASVEMSAVAPPASNGNLLYIGFRGL
                DGVAHVTFPDSGVVCGRMACSPTHDEHLHGQPPEDGPDPLAKILEEVRIANFRFQ
                DESRAVCSWKFACVYDRCLIMAFSVFTICTIGILMSAPNVEAVSKDFA"
        176..1609
        mat_peptide

```

```

BASE COUNT      441 a      601 c      573 g      472 t
ORIGIN

Query Match      81.1%; Score 30; DB 10; Length 2087;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CAGCATATTCAGAGTTCCTGCTACATCG 37
        |||||||
Db      531 CAGCATATTCAGAGTTCCTGCTACATCG 560

RESULT 3
        HSNACHRA7      1509 bp      mRNA      PRI      22-JAN-1998
LOCUS      H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit
DEFINITION      precursor.
ACCESSION      Y08420
VERSION      Y08420.1 GI:2808623
KEYWORDS      nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 1509)
AUTHORS      Groot Kormelink, P.J. and Luyten, W.H.
TITLE      Cloning and sequence of full-length cDNAs encoding the human
                neuronal nicotinic acetylcholine receptor (nAChR) subunits betas
                2 and beta4 and expression of seven nAChR subunits in the human
                neuroblastoma cell line SH-SY5y and/or IMR-32
JOURNAL      FEBS Lett. 400 (3), 309-314 (1997)
MEDLINE      97162233
REFERENCE      2 (bases 1 to 1509)
AUTHORS      Groot Kormelink, P.J.
TITLE      Direct Submission
JOURNAL      Submitted (27-SEP-1996) P.J. Groot Kormelink, Janssen Research
                Foundation, Exp. Mol. Biol. Dept., Turnhoutseweg 30, B-2240 Beerse,
                Belgium
REMARK      Revised by author 22-JAN-1998
COMMENT      On Jan 25, 1998 this sequence version replaced gi:1702915.
FEATURES
        source
        1..1509
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /cell_type="neuroblastoma"
        /cell_line="IMR-32"
        /cell_line="SH-SY5y"
        1..66
        /gene="nAChRA7"
        1..1509
        /gene="nAChRA7"
        /codon_start=1
        /product="nicotinic acetylcholine receptor alpha7 subunit
                precursor"
        /protein_id="CA69697.1"
        /db_xref="PID:e1246208"
        /db_xref="PID:g2808624"
        /db_xref="GI:2808624"
        /translation="MRCSPGGVWGLAASLLHVSLOGEFOKRLKELYNPNLEPVP
                ANDSPLVIFSLIQIMDYDEKNQVLTNTIMWLSQMTDHYLQMNVSSEYGVKTVRF
                PDQIWKPDILLNSADERFDRFTNNVNVNSGHCQYLPGLFKSSCYIDVRFEPD
                VOCHKIKGWSYSGWSLDQMOEADISGITPNGBMDLVGIPGRKSERFECCKEYFP
                DVETFTVRRRTIYGLNLIPCVLISALALVFLPADSGEKISGITVLLSLVEM
                LVAETMPATSDSVPLAIFYASTNIIVGLSVVTVIVLQYHHDPDGCKMPKWTWVI
                LLMNCAMFLRMKRPEDKVRPAQCKORCSLASVEMSAVAPPASNGNLLYIGFRGL
                DGVAHVTFPDSGVVCGRMACSPTHDEHLHGQPPEDGPDPLAKILEEVRIANFRFQ
                DESRAVCSWKFACVYDRCLIMAFSVFTICTIGILMSAPNVEAVSKDFA"
        1..1509
        /gene="nAChRA7"
        /gene="nAChRA7"
        67..1506
        /gene="nAChRA7"
        mat_peptide

```


BASE COUNT 297 a 451 c 429 g 332 t
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1509;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CAGCATATTCAGAGTCTGCTACATCG 37
Db 428 CAGCATATTCAGAGTCTGCTACATCG 457

RESULT 4
LOCUS HSU40583 1977 bp mRNA PRI 19-DEC-1995
DEFINITION Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA,
complete cds.
ACCESSION U40583
VERSION G1125076
GI:1125076

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1977)
Logel, J., Drebing, C., Barnhart, M., Antle, C. and Leonard, S.
Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal
Nicotinic Acetylcholine Receptor in Human Postmortem Brain
Unpublished
2 (bases 1 to 1977)
Leonard, S.
Direct Submission
Submitted (13-NOV-1995) Sherry Leonard, University of Colorado
Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave,
Denver, CO 80262, USA

FEATURES
source Location/Qualifiers
1. 1977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PBshalpha7"
/clone_lib="Clontech lambda gt10 cDNA library: human
hippocampus"
/sex="male"
/tissue_type="hippocampus"
/dev_stage="20 yr old adult"
1. 7
8. 73
8. 1516
/codon_start=1
/product="alpha 7 neuronal nicotinic acetylcholine
receptor"
/protein_id="AA83561.1"
/db_xref="PID:G1125077"
/db_xref="GI:1125077"
/translation="MRCSPGVWMLAASLHVSIOGEFORKLYKELYKNPNLEPV
ANDSQPLTVFSLNLIQIMVDERKNVLTINIMLOMSWTBHYLIQMNSEYPGVTVF
PDGQIMKPDILLYNSADERDAPFTNVLVNSGHCYLPDGLFKSSCYIDVRFPD
VORCKLFGWSYSGWSDLOMOEADISGYIPNEMDLVGIPIGRSRSEYECCEKPP
DVEFTVMRRRTLYGILNLIPICVLISALALVFLPADSGEKISLIGITVLSTVFM
LVAELMPPASDSVPLAQYFASMTIIVGLSVVTVTVLYOHHDHDPGGKMPKTRVI
LNMCAFLMKRPRGEDKVRACQHORCSLASVENSAYAPPAASGNLLYIGFRGL
DGHCYPTDPSGVVCGRMACSPHDEHLHGGCPEDDPLAKTLEVRITANFRQ
DESEAVCSWKFAACVDRCLMAFSVFTIICITGILMSAPNFEAVSKDFA"
1514. 1977
3'UTR polyA-site
BASE COUNT 426 a 567 c 524 g 460 t
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1977;
Best Local Similarity 100.0%; Pred. No. 0.0052;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CAGCATATTCAGAGTCTGCTACATCG 37
Db 435 CAGCATATTCAGAGTCTGCTACATCG 464

RESULT 5
LOCUS HSU62436 1876 bp mRNA PRI 11-JAN-1997
DEFINITION Human nicotinic acetylcholine receptor alpha7 subunit precursor,
mRNA, complete cds.
ACCESSION U62436
VERSION G1458119
GI:1458119

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1876)
Elliot, K.J., Ellis, S.B., Berckhan, K.J., Urrutia, A.,
Chavez-Noriega, L.E., Johnson, E.C., Vellelebi, G. and Harpold, M.M.
Comparative structure of human neuronal alpha 2-alpha 7 and beta
2-beta 4 nicotinic acetylcholine receptor subunits and functional
expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and
beta 4 subunits
J. Mol. Neurosci. 7 (3), 217-228 (1996)
97062879
2 (bases 1 to 1876)
Elliot, K.J.
Direct Submission
Submitted (28-JUN-1996) Kathryn J. Elliot, SIBIA Neurosciences,
Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA

FEATURES
source Location/Qualifiers
1. 1876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="K6alpha7.3"
/clone_lib="SIBIA lambda gt11 library #3 (M. Williams)"
/cell_line="IMR32"
/cell_type="neuroblastoma"
1. 72
73. 138
73. 1581
/codon_start=1
/product="nicotinic acetylcholine receptor alpha7 subunit
precursor"
/protein_id="AAB40114.1"
/db_xref="PID:G1458120"
/db_xref="GI:1458120"
/translation="MRCSPGVWMLAASLHVSIOGEFORKLYKELYKNPNLEPV
ANDSQPLTVFSLNLIQIMVDERKNVLTINIMLOMSWTBHYLIQMNSEYPGVTVF
PDGQIMKPDILLYNSADERDAPFTNVLVNSGHCYLPDGLFKSSCYIDVRFPD
VORCKLFGWSYSGWSDLOMOEADISGYIPNEMDLVGIPIGRSRSEYECCEKPP
DVEFTVMRRRTLYGILNLIPICVLISALALVFLPADSGEKISLIGITVLSTVFM
LVAELMPPASDSVPLAQYFASMTIIVGLSVVTVTVLYOHHDHDPGGKMPKTRVI
LNMCAFLMKRPRGEDKVRACQHORCSLASVENSAYAPPAASGNLLYIGFRGL
DGHCYPTDPSGVVCGRMACSPHDEHLHGGCPEDDPLAKTLEVRITANFRQ
DESEAVCSWKFAACVDRCLMAFSVFTIICITGILMSAPNFEAVSKDFA"
139. 1578
3'UTR
BASE COUNT 369 a 553 c 531 g 423 t
ORIGIN

Query Match 81.1%; Score 30; DB 10; Length 1876;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CAGCATATTCAGAGTCTGCTACATCG 37
|||||

Db 500 CAGGCATATTCAGAGTTCCTGCTACATCG 529

RESULT 6
AF029838 641 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION AF029838
ACCESSION AF029838
NID 93757793
VERSION AF029838.1 GI:3757793
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 641)
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL 9900837
MEDLINE
REFERENCE
AUTHORS 2 (bases 1 to 641)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR1"
/note="RACE product A/B/C/D"
misc_feature <1..>641
/note="alpha 7 neuronal nicotinic receptor mRNA; used in
alternative splicing" 192 c 163 g 135 t

BASE COUNT 151 a 192 c 163 g 135 t

ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 641;
Best Local Similarity 100.0%; Pred. No. 0.0048; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37
|||||
Db 612 CAGGCATATTCAGAGTTCCTGCTACATCG 641

RESULT 7
AF029839 689 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION AF029839
ACCESSION AF029839
NID 93757794
VERSION AF029839.1 GI:3757794
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 689)
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL 9900837
MEDLINE
REFERENCE
AUTHORS 2 (bases 1 to 689)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
misc_feature <1..>689
/note="alpha 7 neuronal nicotinic receptor mRNA; used in
alternative splicing" 210 c 196 g 141 t

BASE COUNT 142 a 210 c 196 g 141 t

ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 689;
Best Local Similarity 100.0%; Pred. No. 0.0049; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37
|||||
Db 660 CAGGCATATTCAGAGTTCCTGCTACATCG 689

RESULT 8
AF036903 1712 bp mRNA PRI 16-OCT-1998
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION AF036903
ACCESSION AF036903
NID 93757807
VERSION AF036903.1 GI:3757807
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 1712)
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene (CHRNA7)
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL 9900837
MEDLINE
REFERENCE
AUTHORS 2 (bases 1 to 1712)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..1712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:W52861"
/db_xref="dbEST:AA037389"
/chromosome="15"
/clone="13p1242"
/clone="11p="Soares papathyroid tumor NHPA library,
Research Genetics/IMAGE Consortium, LBNL"
misc_feature <1..>1712
/note="alpha-7 neuronal nicotinic acetylcholine receptor;
alternatively spliced" 489 c 457 g 410 t

BASE COUNT 356 a 489 c 457 g 410 t

ORIGIN

REFERENCE
AUTHORS 2 (bases 1 to 689)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
misc_feature <1..>689
/note="alpha 7 neuronal nicotinic receptor mRNA; used in
alternative splicing" 210 c 196 g 141 t

BASE COUNT 142 a 210 c 196 g 141 t

ORIGIN

Query Match 81.1%; Score 30; DB 11; Length 689;
Best Local Similarity 100.0%; Pred. No. 0.0049; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTTCCTGCTACATCG 37
|||||
Db 660 CAGGCATATTCAGAGTTCCTGCTACATCG 689

RESULT 8
AF036903 1712 bp mRNA PRI 16-OCT-1998
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION AF036903
ACCESSION AF036903
NID 93757807
VERSION AF036903.1 GI:3757807
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 1712)
Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene (CHRNA7)
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL 9900837
MEDLINE
REFERENCE
AUTHORS 2 (bases 1 to 1712)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..1712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:W52861"
/db_xref="dbEST:AA037389"
/chromosome="15"
/clone="13p1242"
/clone="11p="Soares papathyroid tumor NHPA library,
Research Genetics/IMAGE Consortium, LBNL"
misc_feature <1..>1712
/note="alpha-7 neuronal nicotinic acetylcholine receptor;
alternatively spliced" 489 c 457 g 410 t

BASE COUNT 356 a 489 c 457 g 410 t

ORIGIN

Query Match
Best Local Similarity 81.1%; Score 30; DB 11; Length 1712;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGATATTCAGAGTCTCCTGCTACATCG 37
DB 172 CAGGATATTCAGAGTCTCCTGCTACATCG 201
|||||

RESULT 9
AF037646 1836 bp RNA PRI 16-OCT-1998
LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION
ACCESSION AF037646
NID 93757808
VERSION AF037646.1 GI:3757808
KEYWORDS human.
-ORFCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
Bulling, K., Davis, A., Breese, C., Freedman, R., and Leonard, S.
TITLE Genomic organization and partial duplication of the human alpha7
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 1836)
AUTHORS Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J., and Freedman, R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
FEATURES
SOURCE Location/Qualifiers
1. 1836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:W03952"
/db_xref="dbEST:N73891"
/chromosome="15"
/clone="297134"
/note="alpha-7 scAres fetal liver spleen library, Research
Genetics/IMAGE Consortium, LNL"
/note="unprocessed mRNA with intron"
<1. 1836
/note="alpha-7 neuronal nicotinic acetylcholine receptor
precursor; intron included in 5' region"

BASE COUNT 383 a 503 c 469 g 481 t

misc-feature

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGATATTCAGAGTCTCCTGCTACATCG 37
DB 294 CAGGATATTCAGAGTCTCCTGCTACATCG 323
|||||

RESULT 10
MUSNARS 1848 bp mRNA ROD 02-FEB-1999
LOCUS Mus musculus neuronal nicotinic acetylcholine receptor subunit
DEFINITION
ACCESSION L37663
NID 9790853
VERSION L37663.1 GI:790853
KEYWORDS neuronal acetylcholine receptor subunit alpha 7 unit.

SOURCE Mus musculus (strain BALB/c, sub-species domesticus) neonatal brain
CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Orr-Urtreger, A., Seidlin, M.F., Baldini, A., and Beaudet, A.L.
TITLE Cloning and mapping of the mouse alpha 7-neuronal nicotinic
JOURNAL acetylcholine receptor
MEDLINE Genomics 26 (2), 399-402 (1995)
FEATURES 95324936
SOURCE Location/Qualifiers
1. 1848
/organism="Mus musculus"
/strain="BALB/c"
/sub-species="domesticus"
/db_xref="taxon:10090"
/dev_stage="neonatal"
/tissue_type="brain"
1. 50
51. 1559
/codon_start=1
/product="neuronal acetylcholine receptor subunit alpha 7"
/protein_id="AAC42053.1"
/db_xref="PID:9790854"
/db_xref="GI:790854"
/translation="MCGRRGIMVLAALHNSLOEFPORLYKELVKNYPLERPV
ANDSPPLVFEISLLIMDYDEKNVLTNNIMQMSMTHTYLOMNSSEYGVNVRP
PDQIWMFDILVNSADERDPTHTNVLVAAGCOYLPGLRSCYIDVRRPPD
VOCKLKFSGMSYSGMSLDLQMEADISYIPNGEMDMIGPKRNRKPFCEKEP
DVTYTVRRRTTYGTLNLPCLVALLVLEFLPADSGEKSLDTVLSLTVFM
LVAAEFLPDSVPLTAOYEAASMIVGLSVYTVYLVTHHHDPPGGMPLKTRII
LNMCAWFLMKRPGEDKXPACQHRKRRSLASVELSAGAPTSNGNLYIGFRL
EGMHCAPTPDSGYVCGSLACSPYHDELMHGTSPSDPDLAKILEEYRIANFRQ
DESEVISEKRPACVYDRCLMAFSVFTICTIGILMSAPNFEVANSKQFA"
1560. 1848

BASE COUNT 406 a 487 c 500 g 455 t

3' UTR

Query Match
Best Local Similarity 77.3%; Score 28.6; DB 12; Length 1848;
Matches 31; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCACAGAGCATATTCAGAGTCTCCTGCTACATCG 37
DB 473 CCCTCAGGATATTCAGAGTCTCCTGCTACATCG 507
|||||

RESULT 11
HSCRNA7A 1555 bp mRNA PRI 03-MAR-1994
LOCUS HSCRNA7A
DEFINITION H. sapiens CHRNA7 mRNA, 3' end.
ACCESSION Z23141
NID 9457736
VERSION Z23141.1 GI:457736
KEYWORDS alpha7 nicotinic receptor subunit; cholinergic receptor; CHRNA7
gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Chini, B., Raimond, E., Elgoyhen, A.B., Moralli, D., Balzarotti, M. and
Heinemann, S.
TITLE Molecular cloning and chromosomal localization of the human alpha
JOURNAL 7-nicotinic receptor subunit gene (CHRNA7)
MEDLINE Genomics 19 (2), 379-381 (1994)
REFERENCE 94245214
AUTHORS Chini, B.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille,

AUTHORS

gene	42..1550	/gene="nicotinic receptor alpha 7 subunit"
CDS	42..1550	/gene="nicotinic receptor alpha 7 subunit"

19-MAR-1993
 3030 bp mRNA
 nicotinic receptor alpha 7 subunit [rats, brain, mRNA, 3030 nt].
 553987
 553987
 264770
 553987.1 GI:264770
 Rattus sp. brain.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Theria; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 3030)
 (Seguela, P., Madiche, J., Dineley-Willer, K., Dani, J.A. and
 Patrick, J.W. Molecular cloning, functional properties, and distribution of rat
 brain alpha 7: a nicotinic cation channel highly permeable to
 calcium
 J. Neurosci. 13 (2), 596-604 (1993)
 93147931
 Genbank staff at the National Library of Medicine created this
 entry [NCBI gisbseq 124020] from the original journal article.
 This sequence comes from Fig. 1.
 Location/Qualifiers
 1..3030
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 42..1550
 /gene="nicotinic receptor alpha 7 subunit"
 42..1550
 /gene="nicotinic receptor alpha 7 subunit"
 /note="This sequence comes from Fig. 1. Protein sequence

Sun Sep 19 10:26:01 1999

us-08-956-518a-100.rge

Query Match 60.0%; Score 22.2; DB 10; Length 2251;
Best local similarity 77.1%; Pred. No. 12;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 2 CCCACACAGGCATATTCACAGAGTTCTGCTACATC 36
||||| ||||||| ||| |||
Db 1053 CCCACACAGGCATATTCACAAACACCTGTCCATC 1019

Search completed: September 17, 1999, 22:04:23
Job time: 16446 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:34:34 ; Search time 425.19 Seconds

(without alignments)
21.772 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37

Sequence: 1 ACCCACACAGCATATTCAGAGTCTCTCATACG 37

Scoring table: IDENTITY_NNC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	81.1	1876	1	T48239	Neuronal nicotinic
2	30	81.1	1876	1	V12197	Human neuronal nic
3	30	81.1	1876	1	V44687	V2747 variant huma
4	22	59.5	2769	1	T59196	Neuronal alpha-bun
5	21	56.8	3496	1	T48237	Neuronal nicotinic
6	21	56.8	2082	1	T59527	Alpha4 subunit of
7	21	56.8	2082	1	T59528	Alpha4 subunit of
8	21	56.8	2082	1	T59528	Human neuronal nic
9	21	56.8	2082	1	T59528	Tomato pest resist
10	20.8	56.2	51952	1	V26084	Alpha 2 subunit of
11	19.6	53.0	2277	1	Q90387	Neuronal nicotinic
12	19.6	53.0	2664	1	T48235	Human neuronal nic
13	19.2	51.9	3277	1	V12199	Human aspartic pro
14	19.2	51.9	377	1	V28636	Human aspartic pro
15	19.2	51.9	1375	1	V28623	Human aspartic pro
16	19.2	51.9	452	1	V28626	Nucleotide sequenc
17	19.2	51.9	1298	1	V31665	Human napsin A gen
18	19.2	51.9	1645	1	V27037	Human napsin A gen
19	19.2	51.9	1353	1	V27036	Human napsin A CDN
20	19.2	51.9	1910	1	V27038	Human napsin B CDN
21	19.2	51.4	6273	1	V52154	Streptococcus pneu
22	18.8	50.8	812	1	T13155	Turkey prepro-VIP
23	18.8	50.8	2101	1	T59197	Neuronal alpha-bun
24	18.8	50.8	402	1	V28632	Human aspartic pro
25	18.8	50.8	1860	1	X37192	Recombinant pGO-11
26	18.8	50.8	741	1	X37189	Recombinant pGO-9C
27	18.8	50.8	1476	1	X37190	Recombinant pGO-9C
28	18.8	50.8	1125	1	X37191	Recombinant pGO-11
29	18.8	50.8	651	1	X37193	Recombinant pGO-8P
30	18.8	50.8	1386	1	X37194	Recombinant pGO-8C
31	18.8	50.8	110	1	X37147	PCR primer seq. ID
32	18.4	49.7	1794	1	Q33018	Gene encoding the
33	18.4	49.7	377	1	Q61211	Human brain expres
34	18.4	49.7	1429	1	V04237	Arabidopsis C-14 s
35	18.4	49.7	6588	1	V04238	Arabidopsis C-14 s
36	18.4	49.7	2065	1	V29062	BRCA1 modulator pr
37	18.4	49.7	11309	1	V52241	Streptococcus pneu
38	18.2	49.2	11050	1	V74407	R. communis syntha
39	18.2	49.2	1845	1	Q22614	Synthase factor B
40	18.2	49.2	1845	1	Q22614	Synthase factor B
41	18.2	49.2	1672	1	O63896	DUB-1 enhancer/pro
42	18.2	49.2	1559	1	T91855	Modified tissue pl
43	18.2	49.2	1575	1	N82185	Mouse receptor ME2
			10195	1	T85320	

ALIGNMENTS

RESULT 1
T48239
ID T48239 standard; DNA; 1876 BP.
AC T48239;
DT 09-APR-1997 (first entry)
DE Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.
KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS
FT 73..1581
FT /*tag= a
PN WO9641876-A1.
PD 27-DEC-1996.
PF 07-JUN-1996; U09775.
PR 07-JUN-1995; US-484722.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Elliott KJ, Harpold MM;
DR WPI: 97-065463/06.
DR P-PSDB; W09025.
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
PS used in screening to determine the effect of drugs on the receptor
PT Disclousure; Page 71-73; 108pp; English.
CC A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of
CC the human neuronal nicotinic acetylcholine receptor (nAChR). Host
CC cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7
CC nucleic acids, opt. in combination with other alpha and/or beta
CC subunit nucleic acids (see also T48232-38, T48240-41), express
CC recombinant nAChR subunits useful for identifying cpds. that
CC modulate the activity of human nAChRs.
SQ Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;

Query Match 81.1%; Score 30; DB 1; Length 1876;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGGCATATTCAGAGTCTCTCATACG 37
DB 500 CAGGCATATTCAGAGTCTCTCATACG 529

RESULT 2
V12197
ID V12197 standard; CDNA; 1876 BP.
AC V12197;
DT 14-MAY-1998 (first entry)
DE Human neuronal nicotinic acetylcholine receptor alpha-7 subunit CDNA.
KW Human, neuronal nicotinic acetylcholine receptor; alpha-7 subunit;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS
FT 73..1581
FT /*tag= a
FT /*product= "neuronal nicotinic acetylcholine receptor
PN WO9420617-A2.
PD 15-SEP-1994.
PF 08-MAR-1994; U02447.
PR 08-MAR-1993; US-028031.
PA (SALK-) SALK INST BIOTECHNOLOGY IND ASSOC.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Elliott KJ, Ellis SB, Harpold MM;
DR WPI: 94-303024/37.
DR P-PSDB; W44153.
PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -

PT also transformed cells useful for screening cpds. which modulate
 PT activity of the receptor
 PS Claim 8; Page 78-79; 99pp: English.
 CC The present sequence encodes a human neuronal nicotinic acetylcholine
 CC receptor (nAChR) subunit. The cells expressing the alpha and/or beta
 CC nAChR subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal nAChR.
 CC Subunit specific antibodies may be used to monitor the distribution
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes
 SQ Sequence 1876 BP; 369 A; 533 C; 530 G; 423 T;
 81.1%; Score 30; DB 1; Length 1876;
 est Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37
 500 CAGGCATATTCAGAGTCTCTGCTACATCG 529
 RESULT 3
 ID V44687 standard; cDNA; 1590 BP.
 AC V44687;
 DT 09-OCT-1998 (first entry) nAChR coding sequence.
 DE V274F variant human alpha7 nAChR coding sequence; cancer;
 KW Alpha7 nAChR; alpha7 nicotinic acetylcholine receptor subunit; therapy;
 KW neurodegeneration; enzyme dysfunction; affective disorder; disease;
 KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;
 OS schizophrenia; ss.
 FH Homo sapiens.
 FT Key Location/Qualifiers
 FT CDS 9..1517
 FT /tag- a
 FT CDS
 WT09828331-A2.
 PN 02-JUL-1998.
 PD 22-DEC-1997; U23405.
 PF 20-DEC-1996; US-771737.
 PR (ABBO) ABBOTT LAB.
 BR Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,
 Roch U, Sullivan JP, Touma E;
 P-PSDB; W69216.
 DR WPI; 98-377593/32.
 DR P-PSDB; W69216.
 PT Nucleic acid encoding variant of human alpha7 nicotinic
 PT acetylcholine receptor subunit - used to identify modulators of
 PT the receptor, potentially useful for treating neuro-degeneration,
 PT cancer, affective disorders etc.
 PS Claim 14; Fig 2; 44pp: English.
 CC This sequence encodes the V274F variant of human alpha7 nicotinic
 CC acetylcholine receptor (nAChR) subunit and to identify modulators of
 CC the DNA are used to express the protein and to identify modulators of
 CC alpha7 nAChR activity or cytoprotective agents, e.g. antisense
 CC compounds or antagonists that are potentially useful for treating
 CC neurodegeneration, enzyme dysfunction, affective disorders and immune
 CC compounds or antagonists that are potentially useful for treating
 CC dysfunction, such as cancer, post-hepatic neuropathy, diabetic
 CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,
 CC psychosis and schizophrenia. Probes based on the DNA are used to detect
 CC the DNA in usual hybridisation or amplification tests, while monoclonal
 CC antibodies are used to detect the protein for diagnosis (in vitro or by
 CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nAChR,
 CC the protein has about 100-fold greater sensitivity to cholinergic
 CC receptor agonists (nicotine or acetylcholine) and response to these
 CC agonists decays more slowly, but the wild-type inward rectification is
 CC retained.

SQ Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;
 81.1%; Score 30; DB 1; Length 1590;
 Query Match
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37
 436 CAGGCATATTCAGAGTCTCTGCTACATCG 465
 RESULT 4
 ID T59196 standard; cDNA; 2769 BP.
 AC T59196; 17-JUN-1997 (first entry)
 DT 17-JUN-1997 (first entry)
 DE Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
 KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
 KW ligand binding; ion channel; ss.
 OS Gallus sp.
 FH Gallus sp.
 FT Key Location/Qualifiers
 FT cds 71..1513
 FT /tag- a
 FT signal_peptide 71..136
 FT /tag- b
 FT mat_peptide 137..1510
 FT /tag- c
 WT055599709-A.
 PN 04-FEB-1997.
 PD 28-SEP-1989; 413947.
 PF 28-SEP-1989; US-413947.
 PR (SALK) SALK INST BIOLOGICAL STUDIES.
 PA Lindstrom JM, Schoepfer RD;
 PI WPI; 97-118297/11.
 DR P-PSDB; W12368.
 PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
 PT to screen cholinergic agents and other drugs which may affect ligand
 PT binding, ion channel or other activities of the protein.
 PS Claim 1; Fig 2A-B; 18pp: English.
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
 CC an 18-day embryo chick brain cDNA library using a probe (see also
 CC T59198) based on the N-terminal amino acid sequence of chicken
 CC brain ABBP. The probe isolated partial clone pch29-1, which
 CC encoded the N-terminal portion of alpha1. A subclone, pch29-3
 CC (ATCC 40641), was used to rescreen the library, yielding clone
 CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
 CC A probe based on the C-terminal region of pch31-1 was used obtain
 CC clone pch34-1 (ATCC 40639), encoding the C-terminal portion of
 CC alpha1. The cDNA clones can be used as probes to identify further
 CC ABBP subunits, and in the recombinant prodn. of ABBP.
 SQ Sequence 2769 BP; 790 A; 604 C; 571 G;
 59.5%; Score 22; DB 1; Length 2769;
 Query Match
 Best Local Similarity 83.3%; Pred. No. 3.1;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 8 CAGGCATATTCAGAGTCTCTGCTACATCG 37
 498 CAGGCATATTCAGAGTCTCTGCTACATCG 527
 RESULT 5
 ID T48237 standard; DNA; 3496 BP.
 AC T48237;
 DT 09-APR-1997 (first entry)
 DE Neuronal nicotinic acetylcholine receptor alpha-4 subunit DNA.
 KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;
 KW ligand-gated receptor; ds.

OS Homo sapiens. Location/Qualifiers
 FH Key 232..2115 /tag- a
 FT misc_difference 3344 /tag- b
 FT /note- "base 3344 is given as n in the
 FT specification"
 PN M09641876-A1.
 PD 27-DEC-1996.
 PE 07-JUN-1996; U09775.
 PR 07-JUN-1995; US-484722.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KJ, Harpold WM;
 DR WPI; 97-065463/06.
 DR P-PSDB; W09023.
 PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
 used in screening to determine the effect of drugs on the receptor
 disclosure; Page 58-61; 108pp; English.
 CC A DNA sequence (T48237) codes for the alpha-4 subunit (W09023) of
 the human neuronal nicotinic acetylcholine receptor (nAChR). Host
 cells, esp. mammalian cells or amphibian oocytes, carrying alpha-4
 subunit nucleic acids (see also T48232-36, T48238-41), express
 recombinant nAChR subunits useful for identifying cpds. that
 modulate the activity of human nAChRs.
 CC Sequence 3496 BP; 590 A; 1215 C; 1084 G; 606 T;
 SQ

Query Match 56.8%; Score 21; DB 1; Length 3496;
 Best Local Similarity 73.0%; Pred. No. 8.4;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACAGGAGCATTTCAAGAGTCTCTGCTACATCG 37
 Db 685 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 721

RESULT 6
 T59527
 ID T59527 standard; DNA; 2082 BP.
 AC T59527;
 DT 05-OCT-1997 (first entry)
 DE Alpha4 subunit of normal nAChR.
 KW nAChR; mutation; autosomal dominant nocturnal frontal lobe epilepsy;
 ADNFLE; neuronal acetylcholine receptor; ss.
 KM Homo sapiens.
 Key Location/Qualifiers
 FH misc_rna 902..1168 /tag- a
 FT /note- "ion channel"
 FT 1016..1018 /tag- b
 FT /note- "TCC 248 TTC is the ADNFLE-
 FT inducing mutation (see T59528)"
 PN A09656247-A.
 PD 09-JAN-1997.
 PE 28-JUN-1996; 056247.
 PR 28-JUN-1995; AU-003840.
 PA (UYBO-) UNIV BONN.
 PA (WOME-) UNIV MELBOURNE.
 PI Berovic SE, Mulley JC, Phillips HA, Propping PJ;
 PI Scheffer IE, Steinleink, Sutherland GR, Wallace RH;
 DR WPI; 97-100506/10.
 DR P-PSDB; W11824.
 PT CHRNA4 gene encodes neuronal nicotinic acetylcholine receptor
 fragment - used in diagnosis of autosomal dominant nocturnal frontal
 lobe epilepsy
 CC Claim 5; Fig 13; 20pp; English.
 CC The DNA sequence of the normal nAChR is given in T59527,
 CC the DNA sequence of the ADNFLE-inducing mutant is given in T59528.
 CC Mutations at codon 248, which have the effect of replacing
 CC serine by another amino acid (e.g. phenylalanine) in the sixth
 CC subunit of nAChR have been found to be associated with ADNFLE.
 CC The primers given in T59529 and T59530 were used in the
 CC amplification of part of exon 5.
 SQ Sequence 2082 BP; 367 A; 767 C; 592 G; 355 T;

CC serine by another amino acid (e.g. phenylalanine) in the sixth
 CC amino acid of the transmembrane domain 2 (M2) of the alpha4
 CC subunit of nAChR have been found to be associated with ADNFLE.
 CC The primers given in T59529 and T59530 were used in the
 CC amplification of part of exon 5.
 SQ Sequence 2082 BP; 367 A; 767 C; 592 G; 355 T;

Query Match 56.8%; Score 21; DB 1; Length 2082;
 Best Local Similarity 73.0%; Pred. No. 7.6;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACAGGAGCATTTCAAGAGTCTCTGCTACATCG 37
 Db 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 7
 T59528
 ID T59528 standard; DNA; 2082 BP.
 AC T59528;
 DT 05-OCT-1997 (first entry)
 DE Alpha4 subunit of mutant nAChR (TCC 248 TTC).
 KW nAChR; mutation; autosomal dominant nocturnal frontal lobe epilepsy;
 ADNFLE; neuronal acetylcholine receptor; ss.
 KM Homo sapiens.
 Key Location/Qualifiers
 FH misc_rna 902..1168 /tag- a
 FT /note- "ion channel"
 FT 1016..1018 /tag- b
 FT /note- "TCC 248 TTC is the ADNFLE-
 FT inducing mutation (see T59527)"
 PN A09656247-A.
 PD 09-JAN-1997.
 PE 28-JUN-1996; 056247.
 PR 28-JUN-1995; AU-003840.
 PA (UYBO-) UNIV BONN.
 PA (WOME-) UNIV MELBOURNE.
 PI Berovic SE, Mulley JC, Phillips HA, Propping PJ;
 PI Scheffer IE, Steinleink, Sutherland GR, Wallace RH;
 DR WPI; 97-100506/10.
 DR P-PSDB; W11825.
 PT CHRNA4 gene encodes neuronal nicotinic acetylcholine receptor
 fragment - used in diagnosis of autosomal dominant nocturnal frontal
 lobe epilepsy
 CC Claim 6; Fig 13; 20pp; English.
 CC The DNA sequence of the normal nAChR is given in T59527,
 CC the DNA sequence of the ADNFLE-inducing mutant is given in T59528.
 CC Mutations at codon 248, which have the effect of replacing
 CC serine by another amino acid (e.g. phenylalanine) in the sixth
 CC subunit of nAChR have been found to be associated with ADNFLE.
 CC The primers given in T59529 and T59530 were used in the
 CC amplification of part of exon 5.
 SQ Sequence 2082 BP; 367 A; 767 C; 592 G; 356 T;

Query Match 56.8%; Score 21; DB 1; Length 2082;
 Best Local Similarity 73.0%; Pred. No. 7.6;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ACCGACAGGAGCATTTCAAGAGTCTCTGCTACATCG 37
 Db 632 ACTCCCGCGGCATTACAGAGCTCTGACAGATCG 668

RESULT 8
 V12196
 ID V12196 standard; CDNA; 2363 BP.
 AC V12196;

DT 14-MAY-1998 (first entry)
 DE Human neuronal nicotinic acetylcholine receptor alpha-4 subunit cDNA.
 DE Human neuronal nicotinic acetylcholine receptor; alpha-4 subunit;
 KW brain tissue; screening; NACHR; antibody; ds.
 OS Homo sapiens.
 FT key Location/Qualifiers
 FT CDS 173..2056
 FT /tag= a "neuronal nicotinic acetylcholine receptor
 FT /product= alpha-4 subunit"
 FT
 FT W09420617-A2.
 PN 15-SEP-1994.
 PF 08-MAR-1994; U02447.
 PR 08-MAR-1993; US-028031.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PA (SIBI-) SIBI NEUROSCIENCES INC.
 PI Elliott KJ, Ellis SB, Harpold MM;
 PI WPI: 94-303024/37.
 DR P-PSDB; W44152.
 DR Human neuronal nicotinic acetylcholine receptor subunits and DNA
 also transformed cells useful for screening cpds. which modulate
 activity of the receptor.
 PS Claim 4; Page 74-75; 99PP: English.
 CC The present sequence encodes a human neuronal nicotinic acetylcholine
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta
 CC NACHR subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal NACHR.
 CC Subunit specific antibodies may be used to monitor the distribution
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes.
 CC Sequence 2363 BP; 397 A; 861 C; 696 G; 403 T;
 SQ
 Query Match 56.8%; Score 21; DB 1; Length 2363;
 Best Local Similarity 73.0%; Pred. No. 7.8;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 ACCACACAGCAGCATATTCAGAGTCTCTGCTACATCG 37
 DB 626 ACTCCCGCCGCCATTACAGAGCTCTGTCGACATCG 662
 ID V26084/c
 AC V26084; standard: cDNA; 51952 BP.
 DT 07-JUN-1999 (first entry)
 DE Tomato pest resistance M1 gene (copy 1).
 DE Tomato pest resistance; nematode resistance; disease resistance; M1 gene;
 KW tomato; transgenic plant; crop protection; biological control; ss.
 OS Lycopersicon esculentum.
 FT key Location/Qualifiers
 FT promoter 10071..14691
 FT /tag= a
 FT /note= "copy 2 promoter; specifically claimed in
 FT nucleic acid construct of Claim 9"
 FT CDS 16084..19932
 FT /tag= b
 FT /note= "M1 gene (copy 2); encodes W55975"
 FT promoter 42798..44461
 FT /tag= c
 FT /note= "copy 1 gene promoter"
 FT CDS 45102..48944
 FT /tag= d
 FT /note= "M1 gene (copy 1); encodes W55974"
 PN W09815171-A1.

PD 16-APR-1998.
 PF 09-OCT-1997; U18802.
 PR 10-OCT-1996; US-028191.
 PA (REGC) UNIV CALIFORNIA.
 PA Bodeau J, Kaloshian I, Milligan S, Williamson VM,
 PI Yaghoobi J;
 PI WPI: 98-240529/21.
 DR P-PSDB; W55974, W55975.
 DR Nucleic acids encoding M1 polypeptide(s) conferring nematode
 PT resistance - useful to produce transgenic plants resistant to these
 PT and other pests, and in marker-aided selection to assess cultivars
 PT for resistance.
 PS Claim 11; Page 24-39; 55PP: English.
 CC This is the nucleotide sequence of the tomato M1 locus associated
 CC with nematode resistance, obtained from a bacterial artificial
 CC chromosome (BAC). M1 was localised by genetic analysis to a region
 CC of the tomato genome of about 65 kb. DNA corresponding to this
 CC region was cloned into BAC vectors. Sequence analysis of a 52 kb
 CC BAC3 insert identified 3 open reading frames, one of which is
 CC probably a pseudogene. By RNA blot analysis, transcripts of
 CC approximately 4 kb corresponding to copy 1 and copy 2 were found in
 CC both resistant and susceptible tomato roots and in leaves of
 CC resistant but not susceptible plants. cDNA sequences corresponding
 CC to full-length transcripts of copy 1 (see V26082) and copy 2 (see
 CC V26083) were obtained. The encoded polypeptides (see W55974-75)
 CC are 91% identical and contain structural features similar to known
 CC plant resistance genes (R genes) of the nucleotide binding site/
 CC leucine-rich repeat (NBS/LRR) family. A recombinant expression
 CC cassette comprising an M1 polynucleotide and an operably linked
 CC plant promoter can be used to enhance nematode resistance in plants
 CC especially tomatoes (claimed). Transgenic plants can also be
 CC constructed using a M1 promoter with heterologous genes; the M1
 CC temporal and spatial patterns and at similar levels to resistance
 CC genes.
 CC Sequence 51952 BP; 17728 A; 8444 C; 8647 G; 17111 T;
 SQ
 Query Match 56.2%; Score 20.8; DB 1; Length 51952;
 Best Local Similarity 78.1%; Pred. No. 16;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 5 ACACAGCATATTCAGAGTCTCTGCTACATC 36
 DB 30281 ACAAGGCGAGTTCAGAGTCTCTGCTACATC 30250
 ID Q90387
 AC Q90387; standard: cDNA; 2277 BP.
 DT 30-NOV-1995 (first entry)
 DE Alpha 2 subunit of a human neuronal nicotinic acetylcholine receptor;
 KW human NACHR; neuronal nicotinic acetylcholine receptor;
 OS Homo sapiens.
 FT key Location/Qualifiers
 FT CDS 166..1755
 FT /tag= a
 PN W09513299-A.
 PN 18-MAY-1995.
 PF 08-NOV-1994; U12859.
 PR 08-NOV-1993; US-149503.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PA Elliott KJ, Ellis SB, Harpold MM;
 PI WPI: 95-194036/25.
 DR P-PSDB; R73966.
 DR New human neuronal nicotinic acetylcholine receptor alpha2 subunit DNA
 PT used to develop prods. for detection, diagnosis and therapy and for
 PT modulating activity.
 PS Claim 2; Page 43-46; 54PP: English.
 CC DNA encoding the human NACHR alpha2 subunit was isolated from a
 CC human thalamus tissue cDNA library using corresp. rat cDNA. The

CC Insert of one clone cdt. was ligated with the insert of another
 CC clone to generate a full-length alpha 2 subunit cDNA. The DNA can be
 CC used to identify function NACHRs. Cells contg. the DNA can be
 CC for screening to identify cpds. which modulate the activity of human
 CC NACHRs. The human NACHR alpha 2 subunit can be used to product
 CC antibodies which can be used in immunohistochemistry, diagnosis and
 CC therapy. The nucleic acids can be used for analysing disease states
 CC and creating animal models.
 SO Sequence 2277 BP; 451 A; 696 C; 643 G; 487 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2277;
 Best Local Similarity 73.5%; Pred. No. 29;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37
 NB 688 CCCCCGCCATCTACAGAGCTCTGCGACATCG 721

RESULT 11

ID T48235 standard; DNA: 2664 BP.
 AC T48235;
 DT 09-APR-1997 (first entry)
 DE Neuronal nicotinic acetylcholine receptor alpha-2 subunit DNA.
 KW Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT cds
 FT 555..2141
 FT /*tag= a

PN W09641876-A1.
 PD 27-DEC-1996.
 PF 07-JUN-1996; U09775.
 PR 07-JUN-1995; US-484722.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KJ, Harpold WM;
 DR WPI; 97-065463/06.
 DR P-PSDB; W096021.
 PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
 PT used in screening to determine the effect of drugs on the receptor
 PS disclosure; Page 49-52; 108pp; English.
 CC A DNA sequence (T48235) codes for the alpha-2 subunit (W09021) of
 CC the human neuronal nicotinic acetylcholine receptor (nAChR). Host
 CC cells, esp. mammalian cells or amphibian oocytes, carrying alpha-2
 CC nucleic acids, opt. in combination with other alpha and/or beta
 CC subunit nucleic acids (see also T48232-34, T48236-41), express
 CC recombinant nAChR subunits useful for identifying cpds. that
 CC modulate the activity of human NACHRs.
 SO Sequence 2664 BP; 518 A; 815 C; 743 G; 588 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2664;
 Best Local Similarity 73.5%; Pred. No. 30;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37
 DB 1077 CCCCCGCCATCTACAGAGCTCTGCGACATCG 1110

RESULT 12

ID V12199 standard; cDNA: 2277 BP.
 AC V12199;
 DT 14-MAY-1998 (first entry)
 DE Human neuronal nicotinic acetylcholine receptor alpha-2 subunit cDNA.
 KW Human; neuronal nicotinic acetylcholine receptor; alpha-2 subunit.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT CDS
 FT 106..1175

FT /*tag= a
 FT /product= "neuronal nicotinic acetylcholine receptor
 FT alpha-2 subunit"

PN W09420617-A2.
 PD 15-SEP-1994.
 PF 08-MAR-1994; U02447.
 PR 08-MAR-1993; US-028031.
 PA (SALK-) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Elliott KJ, Ellis SB, Harpold WM;
 DR WPI; 94-303024/37.
 DR P-PSDB; W44155.

PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -
 PT also transformed cells useful for screening cpds. which modulate
 PT activity of the receptor
 PS disclosure; Page 67-68; 99pp; English.
 CC The present sequence encodes a human neuronal nicotinic acetylcholine
 CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta
 CC NACHR subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal NACHR.
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes.
 SO Sequence 2277 BP; 451 A; 695 C; 643 G; 486 T;

Query Match 53.0%; Score 19.6; DB 1; Length 2277;
 Best Local Similarity 73.5%; Pred. No. 29;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 4 CACACAGCATATTCAGAGTCTCTGCTACATCG 37
 DB 688 CCCCCGCCATCTACAGAGCTCTGCGACATCG 721

RESULT 13

ID V28636/c
 AC V28636;
 DT 29-JUL-1998 (first entry)
 DE Human aspartic protease partial cDNA EST 857644 SEQ ID NO:14.
 KW Human; aspartic protease; EST; endothelin; serum amyloid A protein;
 KW pro-opiomelanocortin prohormone; ss.
 OS Homo sapiens.
 PN W09811236-A1.
 PD 19-MAR-1998.
 PF 09-SEP-1997; G02426.
 PR 11-SEP-1996; GB-018966.
 PA (SMIK-) SMITHKLINE BEECHAM PLC.
 PA (UYMA-) UNIV MALES.
 PI H11 J, Kay J, Powell D;
 DR WPI; 98-207396/18.

PT Human aspartic protease and related DNA - which may be involved in
 PT processing of endothelin and pro-opiomelanocortin pro-hormone(s)
 PS disclosure; Page 23; 32pp; English.
 CC The present sequence represents a partial cDNA sequence from a human
 CC aspartic protease, isolated from a cDNA library of human origin. The
 CC present invention describes novel human aspartic proteases and also ESTs
 CC from human cDNA libraries having partial DNA sequences which encodes an
 CC aspartic protease. Compounds which inhibit aspartic protease, especially
 CC used in therapy where needed. Important functions of aspartic proteases
 CC are the processing of endothelin and pro-opiomelanocortin prohormones.
 CC They may also be involved in the processing of serum amyloid A protein.
 SO Sequence 377 BP; 62 A; 97 C; 121 G; 89 T;

Query Match 51.9%; Score 19.2; DB 1; Length 377;
 Best Local Similarity 75.0%; Pred. No. 31; Mismatches 8; Indels 0; Gaps 0;
 Matches 24; Conservative 0;

5 ACACAGGATATTCAGAGTCTGCTACATC 36
 179 AGACAGGCTTATCCATAGCCCTGCTCACC 148

RESULT 14
 V28623/C
 ID V28623 standard; cDNA: 1375 BP.

AC 29-JUL-1998 (first entry)
 DE Human aspartic protease encoding cDNA SEQ ID NO:2.
 KM Human; aspartic protease; EST: endothelin; serum amyloid A protein;
 OS pro-opiomelanocortin prohormone; ss.

Location/Qualifiers
 Key 26.1339

/*tag= a
 /product= "aspartic protease"
 /transl_except= (pos:132..1324,aa:Arg)
 /transl_except= (pos:1325..1327,aa:Ala)
 /transl_except= (pos:1328..1330,aa:Ser)
 /transl_except= (pos:1331..1333,aa:Ala)
 /transl_except= (pos:1334..1336,aa:Thr)
 /transl_except= (pos:1337..1339,aa:Gln)
 /note= "no stop codon given; the CDS is marked according to the protein given in the specification but could decode for its full length"

WO9811236-A1.

19-MAR-1998.

09-SEP-1997; G02426.

11-SEP-1996; GB-018966.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PA (UYMA-) UNIV WALES.

PI Hill J, Kay J, Powell D;

DR WPI: 98-207396/18.

P-PSDB: W57042.

PT Human aspartic protease and related DNA - which may be involved in processing of endothelin and pro-opiomelanocortin pro-hormone(s)

PS Claim 2: Page 20; 32pp; English.

CC The present sequence encodes a human aspartic protease. The present invention describes novel human aspartic proteases which encode an

CC human cDNA libraries having partial DNA sequences which inhibit aspartic protease, especially

CC aspartic protease. Compounds which inhibit aspartic protease, especially

CC antibodies can be used in therapy where needed. The protease can also be

CC used in the processing of endothelin and pro-opiomelanocortin prohormones.

CC are the processing of endothelin and pro-opiomelanocortin pro-hormone(s).

CC They may also be involved in the processing of serum amyloid A protein.

CC Sequence 1375 BP; 255 A; 419 C; 396 G; 305 T;

Query Match 51.9%; Score 19.2; DB 1; Length 1375;
 Best Local Similarity 75.0%; Pred. No. 38; Mismatches 8; Indels 0; Gaps 0;
 Matches 24; Conservative 0;

5 ACACAGGATATTCAGAGTCTGCTACATC 36
 674 AGACAGGCTTATCCATAGCCCTGCTCACC 643

RESULT 15

V28626/C
 ID V28626 standard; cDNA: 452 BP.

AC 29-JUL-1998 (first entry)

DE Human aspartic protease partial cDNA EST 424772 SEQ ID NO:4.

KM Human; aspartic protease; EST: endothelin; serum amyloid A protein;

OS pro-opiomelanocortin prohormone; ss.

OS Homo sapiens.

PN WO9811236-A1.
 PD 19-MAR-1998.
 PE 09-SEP-1997; G02426.
 PR 11-SEP-1996; GB-018966.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PA (UYMA-) UNIV WALES.

PI Hill J, Kay J, Powell D;

DR WPI: 98-207396/18.

PT Human aspartic protease and related DNA - which may be involved in processing of endothelin and pro-opiomelanocortin pro-hormone(s)

PS Disclosure: Page 21; 32pp; English.

CC The present sequence represents a partial cDNA sequence from a human

CC aspartic protease, isolated from a cDNA library of human origin. The

CC present invention describes novel human aspartic proteases which encode an

CC from human cDNA libraries having partial DNA sequences which inhibit aspartic

CC aspartic protease. Compounds which inhibit aspartic protease, especially

CC antibodies can be used in therapy where needed. The protease can also be

CC used in the processing of endothelin and pro-opiomelanocortin prohormones.

CC are the processing of endothelin and pro-opiomelanocortin pro-hormone(s).

CC They may also be involved in the processing of serum amyloid A protein.

CC Sequence 452 BP; 81 A; 109 C; 137 G; 114 T;

Query Match 51.9%; Score 19.2; DB 1; Length 452;
 Best Local Similarity 75.0%; Pred. No. 32; Mismatches 8; Indels 0; Gaps 0;
 Matches 24; Conservative 0;

5 ACACAGGATATTCAGAGTCTGCTACATC 36
 211 AGACAGGCTTATCCATAGCCCTGCTCACC 180

Search completed: September 18, 1999, 00:34:37
 Job time: 19037 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:26 : Search time 2825.05 Seconds

(without alignments)
25.834 Million cell updates/sec

Title: US-08-956-518a-100

Perfect score: 37
Sequence: 1 ACCCAGCAGGCAATTCAGAGTCTGCTACATCG 37

Scoring table: IDENTITY_NUC

arched: 2546578 seqs, 986266752 residues

database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	37	100.0	261	21	F08364	F08364 HSCZSA11 n
2	37	100.0	502	39	AA889613	AA889613 ak28e11.s
3	37	100.0	497	43	AI241498	AI241498 qh69c08.x
4	37	100.0	476	45	AI368029	AI368029 q945a03.x
5	36	97.3	421	25	N71634	N71634 za1c07.s1
6	29	78.4	405	26	W93369	W93369 zh46a03.s1
7	27	73.0	106	26	W86805	W86805 zh46g04.s1
8	26.2	70.8	560	26	W52861	W52861 zc03e02.r1
9	22.8	61.6	293	23	R88873	R88873 ym97b10.r1
10	22.4	60.5	511	47	AI509992	AI509992 m119h09.y
11	22.2	60.0	426	22	H01869	H01869 y132a12.r1
12	22.2	60.0	411	22	R32633	R32633 yh73h02.r1
13	22.2	60.0	393	22	R71355	R71355 y154f12.r1
14	22	59.5	418	36	AA621981	AA621981 nq24c12.s
15	21.8	58.9	395	46	AA664815	AA664815 U1-R-CO-g
16	21.2	57.3	416	22	R82535	R82535 y119b06.r1
17	21.2	57.3	409	25	W03952	W03952 za62c08.r1
18	20.8	56.2	1161	26	W41959	W41959 mce6b10.r1
19	20.8	56.2	728	50	AA050499	AA050499 m120f02.r
20	20.6	55.7	337	20	AI683454	AI683454 tw51f01.x
21	20.6	55.7	337	20	AA826609	AA826609 o134d07.s
22	20.6	55.7	337	20	AA826609	AA826609 o134d07.s
23	20.2	54.6	398	40	AA9929019	AA9929019 o04d04.s
24	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
25	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
26	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
27	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
28	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
29	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
30	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
31	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
32	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
33	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
34	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
35	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
36	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
37	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
38	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
39	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
40	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
41	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
42	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
43	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
44	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
45	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
46	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
47	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
48	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
49	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
50	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
51	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
52	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s
53	20.2	54.6	373	40	AA9929019	AA9929019 o04d04.s

ALIGNMENTS

RESULT 1
F08364/c
LOCUS F08364
DEFINITION HSCZSA11 normalised infant brain cDNA Homo sapiens
ACCESSION F08364
NID 9676930
VERSION F08364.1 GI:676930

261 bp mRNA
c-zsa11, mRNA sequence.

EST 22-FEB-1995
c-zsa11, mRNA sequence.

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 261)
AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
Devignes, M.D., Duprat, S., Houligate, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y.,
Sebastiani, Kaakchhis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534

JOURNAL
MEDLINE
COMMENT

TITLE
Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: y1c-zsall
Insert length: 664 Std Error: 0.00
Seq primer: (-21)M13 universal
High quality sequence stop: 436.
Location/Qualifiers

FEATURES
source
1..261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zsall"
/clone_lib="normalized infant brain cDNA"
/sex="female"
/tissue="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site: 1: HindIII;
Site: 2: NotI; sex: female; dev_stage=3 months old;
Isolate-muscular atrophy patient; tissue-type-total
brain; total mRNA was oligo-(dt) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
laamid BA vector. Clone library from B. Soares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press

BASE COUNT
71 a 65 c 60 g 63 t 2 others

ORIGIN
Query Match 100.0%; Score 37; DB 21; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAGCATATTCAGAGTCCGCTACATCG 37
|||||
Db 253 ACCCACACAGCATATTCAGAGTCCGCTACATCG 217

RESULT 2
AA889613 502 bp mRNA EST 04-JAN-1999
LOCUS AA889613
DEFINITION ak28611.s1 Soares, testis_NHT Homo sapiens cDNA clone IMAGE:1407308
3' similar to gb:70297 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN,
ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION AA889613
NID 93016492
VERSION AA889613.1 GI:3016492
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 502)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced gi:936210.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/brp/image/image.html

FEATURES
source
Insert length: 376 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 255.
Location/Qualifiers

1..502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1407308"
/clone_lib="Soares-testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker. Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5']
TGTACCATCATGAGTGGAGCGCCGCCCATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to cots, and was
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT
128 a 98 c 122 g 154 t

ORIGIN
Query Match 100.0%; Score 37; DB 39; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCACACAGCATATTCAGAGTCCGCTACATCG 37
|||||
Db 319 ACCCACACAGCATATTCAGAGTCCGCTACATCG 355

RESULT 3
A1241498 497 bp mRNA EST 04-NOV-1998
LOCUS A1241498
DEFINITION q169d08.x1 Soares, fetal_liver_spleen_INFIS.S1 Homo sapiens cDNA
clone IMAGE:1849935 3' similar to gb:70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION A1241498
NID 93636895
VERSION A1241498.1 GI:3836895
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 497)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Aug 21, 1998 this sequence version replaced.
Contact: Robert Strausberg, Ph.D.

15' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

110 a 83 c 95 g 125 t 8 others

Query Match 97.3%; Score 36; DB 25; Length 421;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAGGCAATTCAGAGTTCCTGCTACATC 36
DB 320 ACCCAGACAGGCAATTCAGAGTTCCTGCTACATC 355

DEFINITION
W93369 405 bp mRNA EST 16-JUL-1996
zh46a03.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:415084 3' similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN); mRNA sequence.

ACCESSION
W93369
VERSION
W93369.1 GI:1422491
KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 405)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matis, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE
JOURNAL
COMMENT
The WashU-Merck EST Project
Unpublished (1995)
On Sep 21, 1992 this sequence version replaced gi:279006.

FEATURES
SOURCE
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoImage.lnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 247.
Location/Qualifiers
1..405
/organism="Homo sapiens"
/db_xref="GDB:1333554"
/db_xref="taxon:9606"
/map="19p12-p13.1"
/clone="IMAGE:415084"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

Query Match 78.4%; Score 29; DB 26; Length 405;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCAGACAGGCAATTCAGAGTTCCTG 29
DB 312 ACCCAGACAGGCAATTCAGAGTTCCTG 340

DEFINITION
W86805 106 bp mRNA EST 02-FEB-1997
zh64604.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:416886 3' similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN); mRNA sequence.

ACCESSION
W86805
VERSION
W86805.1 GI:1400553
KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 106)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Matis, M.

TITLE
JOURNAL
COMMENT
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On May 9, 1995 this sequence version replaced gi:802594.

FEATURES
SOURCE
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoImage.lnl.gov) for further information.
Trace considered overall poor quality
Insert length: 1097 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 1.
Location/Qualifiers
1..106
/organism="Homo sapiens"
/db_xref="GDB:1325356"
/db_xref="taxon:9606"
/clone="IMAGE:416886"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library

TGTTACCAACTGTAGATGGAGGAGCGCCGCCACAACTTTTTTTTTTTTTTTTTTTTTT
7-3'-), double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT77 vector (Pharmacia). Library went through one round of normalization to a Cot - 5 library constructed by Benzo Soares and M Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M. Fatima Bonalato. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a

raptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT 49 a 90 c 80 g 69 t 5 others
ORIGIN

Query Match 61.6%; Score 22.8; DB 23; Length 293;
Best Local Similarity 77.1%; Pred. No. 8.1;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CCCACAGGCATATTCAGAGTCTGCTGATAC 36
Db 118 CCCACAGGCATATTCAGAGTCTGCTGATC 152

---SULT 10
09992

DEFINITION A1509992 511 bp mRNA EST 12-MAR-1999
m19h09.y1 Soares mouse embryo NDM3.5 14.5 Mus musculus cDNA
clone IMAGE:476609 5' similar to gb:U15647_cds1 Mus musculus
(MOUSE);, mRNA sequence.

ACCESSION A1509992

NID 94408897

VERSION A1509992.1 GI:4408897

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 511)

AUTHORS

Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
On Jun 5, 1998 this sequence version replaced gi:3188684.

TITLE JOURNAL

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Putative full length read
vector to vector length is 577
Seq primer: -40bp from Glibco
High quality sequence stop: 447.
Location/Qualifiers

FEATURES
source

1..511
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="476609"
/clone_lib="Soares mouse embryo NDM3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p773p-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo (dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGGCAATATTTTTTTTTTTTTTTTTT

BASE COUNT 186 a 135 c 76 g 113 t 1 others
ORIGIN

Query Match 60.5%; Score 22.4; DB 47; Length 511;
Best Local Similarity 81.2%; Pred. No. 12;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CCACAGGCATATTCAGAGTCTGCTGATCA 34
Db 443 CCAACTGTGATCTTCAAGAGTCTCTACTACA 474

RESULT 11

H01869

LOCUS

DEFINITION

H01869 426 bp mRNA EST 19-JUN-1995
Y132a12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:150430 5', mRNA sequence.

ACCESSION

NID 9664802

VERSION H01869.1 GI:864802

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 426)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F.,
Treviskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On May 9, 1995 this sequence version replaced gi:804305.

TITLE JOURNAL

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Insert Size: 1235

High quality sequence stops: 322

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1235 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 322.
Location/Qualifiers

FEATURES
source

1..426
/organism="Homo sapiens"
/db_xref="GDB:562415"
/db_xref="taxon:9606"
/clone_image="150430"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta. Vector: p773p (Pharmacia) with a
modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5'
AATCGAAGATTCGCGCGGCGGAGGAAATTTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 69 a 136 c 122 g 95 t 4 others

ORIGIN

Query Match 60.0%; Score 22.2; DB 22; Length 426;
Best Local Similarity 77.1%; Pred. No. 15;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 36
Db 33 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 67

RESULT 12

LOCUS R32633 411 bp mRNA EST 28-APR-1995
DEFINITION YN73H02.r1 Soares Placenta Nb2HP Homo sapiens cDNA clone
IMAGE:135411 5', mRNA sequence.

ACCESSION R32633
NID 9788476
VERSION R32633.1 GI:788476
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1281
High quality sequence stops: 244
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1281 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 244.
Location/Qualifiers

FEATURES

source

1..411
/organism="Homo sapiens"
/db_xref="GDB:541344"
/db_xref="taxon:9606"
/clone="IMAGE:135411"
/clone.lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGCGAAGATTCGCGCGCGAGATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 74 a 123 c 117 g 94 t 3 others

ORIGIN

Query Match 60.0%; Score 22.2; DB 22; Length 411;
Best Local Similarity 77.1%; Pred. No. 15;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 36
Db 33 CCCACAGAGCATATTCAGAGTTCCTGCTACATC 67

RESULT 13

LOCUS R71355 393 bp mRNA EST 01-JUN-1995
DEFINITION YL54F12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:143087 5', mRNA sequence.

ACCESSION R71355
NID 9844872
VERSION R71355.1 GI:844872
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 18, 1995 this sequence version replaced gi:811226.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1209
High quality sequence stops: 347
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1209 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 347.
Location/Qualifiers

FEATURES

source

1..393
/organism="Homo sapiens"
/db_xref="GDB:552210"
/db_xref="taxon:9606"
/map="17"
/clone="IMAGE:143087"
/clone.lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGCGAAGATTCGCGCGCGAGATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bernaldo. "

BASE COUNT 68 a 129 c 113 g 79 t 4 others

Query Match 60.0%; Score 22.2; DB 22; Length 393;
 Best Local Similarity 77.1%; Pred. No. 15;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

2 CCCACACAGGCATATTCAGAGTTCCTCTACATC 36
 111 111 111 111 111 111 111
 Db 32 CCCACACCCCATATTCATGAGTACTCTGGATC 66

RESULT 14
 AA621981/c

LOCUS 418 bp mRNA
 DEFINITION hg24c12.s1 NCI-CGAP-Co10 Homo sapiens CDNA clone IMAGE:1144822.3
 similar to TR:G1339923 G1339923 PHENOL SULFOTRANSFERASE SUBUNIT ;
 mRNA sequence.

ACCESSION AA621981
 NID 92525857
 VERSION AA621981.1 GI:2525857
 KEYWORDS EST.
 SOURCE human

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 418)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Sep 12, 1996 this sequence version replaced gi:1407520.

REFERENCE 1 (bases 1 to 418)
 AUTHORS NCI-CGAP
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1407520.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdip/image/image.html

Trace considered overall poor quality
 Insert length: 1697 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 1.

FEATURES

source

1. 418
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1144822"
 /clone_id="NCI-CGAP-Co10"
 /tissue_type="Colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER colon tumor; and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo (N-Soares4)."

BASE COUNT 107 a 82 c 116 g 113 t

ORIGIN

Query Match 59.5%; Score 22; DB 36; Length 418;
 Best Local Similarity 83.3%; Pred. No. 18;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 1 ACCACACAGGCATATTCAGAGTTCCTCTGC 30
 111 111 111 111 111 111 111
 Db 99 ACCGACACGTCATTCACAGATCCCTGC 70

RESULT 15
 AA964815/c
 LOCUS 395 bp mRNA
 DEFINITION UI-R-CO-gx-a-10-0-UI.s1 UI-R-CO Rattus norvegicus CDNA clone
 UI-R-CO-gx-a-10-0-UI 3', mRNA sequence.

ACCESSION AA964815
 NID 94279689
 VERSION AA964815.1 GI:4279689
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 395)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704447
 On May 18, 1998 this sequence version replaced gi:3138307.

REFERENCE 1 (bases 1 to 395)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704447
 COMMENT On May 18, 1998 this sequence version replaced gi:3138307.

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult kidney library. cDNA library preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics The following repetitive elements were found in this cDNA
 sequence: 1-54, >AT richlow_complexity 57-143,
 >(CANA)n#Simple_repeat
 Seq primer: M13 forward.

FEATURES

source

1. 395
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /map_6p21.3: 5q33
 /clone_id="UI-R-CO-gx-a-10-0-UI"
 /clone_lib="UI-R-CO"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CO
 library is a subtracted library derived from the UI-R-A1
 and UI-R-E1 libraries. The UI-R-A1 library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1
 library consisted of a mixture of individually tagged
 normalized libraries constructed from 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which
 allows identification of the library of origin of a clone
 within the mixture. The subtracted library (UI-R-CO) was
 constructed as follows: PCR amplified cDNA inserts from a
 pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
 been derived was used as a driver in a hybridization with
 the pooled UI-R-A1 and UI-R-E1 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-CO
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)."

BASE COUNT 71 a 76 c 108 g 140 t
ORIGIN

Query Match 58.9%; Score 21.8; DB 46; Length 395;
Best Local Similarity 78.8; Pred. No. 21;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACACAGGCATATTCAGAGTTCCTGCTACATC 36
||||| | | ||||| ||||| |||
Db 223 CACACAGCCAGACAGACAGATCCTGCCACAGC 191

Search completed: September 17, 1999, 21:28:29
Job time: 14293 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:04:23 ; Search time 1962.17 Seconds
(without alignments)
635.358 Million cell updates/sec

Title: US-08-956-518a-101

Sequence: 392

1 AGAAGCAAGAGAGAGTAG.....CTGACGCTCCGGACTCAAC 392

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pl3: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_to: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vl: *
18: em_fun: *
19: em_htg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vl: *
34: gb_htg2: *
35: gb_htg1: *
36: gb_in1: *
37: gb_in2: *
38: em_bal: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pr4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	392	100.0	392	11	AF029837	Homo sapi

2	72	18.4	1876	10	HS62436	U62436 Human nicot
3	71.8	18.3	168819	35	AC007686	AC007686 Homo sapi
4	70	17.9	342379	34	CEY3986	295399 Caenorhabdi
5	69.8	17.8	2087	10	HSARA7A	X70297 H. sapiens m
6	68	17.3	1072	17	HS2IEA	M29384 Herpes simp
7	65.8	16.8	1560	17	HS2IE	M29384 Herpes simp
8	63.2	16.1	43058	10	HMG81	284721 Human DNA s
9	63.2	16.1	2885	10	HOMBA3	J00184 Human alpha
10	63	16.1	154746	17	HSV2HG52	286099 Herpes simp
11	61.6	15.7	14936	10	CH19R30879	AD000684 Homo sapi
12	61.6	15.7	14440	10	HSUSP2	Y07661 H. sapiens U
13	61.6	15.7	14440	10	HSUSP2	Y07661 H. sapiens U
14	61.4	15.7	38532	1	SCB94	AL049628 streptomy
15	61	15.6	1927	3	CEGPEPPAM	Z12168 C. familiar
16	61	15.6	557	4	XELRGE12	K01371 X. laevis oo
17	61	15.6	3924	4	XELRGE12	J00999 X. laevis ex
18	61	15.6	7634	4	XL28SR	X59734 X. laevis 28
19	61	15.6	8153	4	XLRN01	X02995 Xenopus lae
20	61	15.6	248287	35	AC007225	AC007225 Homo sapi
21	59	15.1	3252	10	HSTAF113	Y11354 H. sapiens m
22	59	15.1	634	14	G40486	G40486 26971 Zebra
23	58.8	15.0	210636	11	AC006443	AC006443 Homo sapi
24	57.8	14.7	179436	35	AC009066	AC009066 Mus muscu
25	57.6	14.7	181884	34	HS35C18	AL022327 Homo sapi
26	57	14.5	267925	35	AC007298	AC007298 Homo sapi
27	56.8	14.5	204263	35	AC007224	AC007224 Homo sapi
28	56.4	14.4	152261	17	HE1CG	X14112 Herpes simp
29	56.4	14.4	152261	17	HE1CG	X14112 Herpes simp
30	56.4	14.4	6633	17	HEHSY1G3	X06461 Herpes simp
31	56.4	14.4	26245	17	HS1US	L00036 HSV1 (strat
32	56.4	14.4	26245	17	HS1US	L00036 HSV1 (strat
33	56.2	14.3	24800	1	SC9B1	AL049727 Streptomy
34	56.2	14.3	1766	17	HS1S17	R01835 HSV-1 (stra
35	55.8	14.2	37321	11	AC005498	AC005498 Homo sapi
36	55.6	14.2	168588	35	AC007613	AC007613 Homo sapi
37	55.6	14.2	247297	35	AC007616	AC007616 Homo sapi
38	55.2	14.1	4352	12	MMUSF2S1	U12282 Mus musculi
39	54.8	14.0	10144	10	HUMHNG17	L17131 Human high
40	54.6	13.9	27874	12	MMU71085	U71085 Mus musculu
41	54.6	13.9	110511	17	EVCCGA	AL049611 Homo sapi
42	54.2	13.8	7861	34	EVCCGA	X74312 Encephalomy
43	53.8	13.7	9701	9	AB01733S3	AB017335 Homo sapi
44	53.8	13.7	42965	34	HS367G8	297634 Homo sapien
45	53.8	13.7	240229	35	AC003059	AC003059 Mus muscu

ALIGNMENTS

RESULT	1	AF029837	392 bp	DNA	PRI	16-DEC-1998
LOCUS	AF029837					
DEFINITION	Homo sapiens alpha 7 neuronal nicotinic receptor gene, promoter					
ACCESSION	AF029837					
NID	93757792					
VERSION	AF029837.1	GI:3757792				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S., genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene					
TITLE	JOURNAL MEDLINE 99000837					
REFERENCE	2 (pages 1 to 392)					
AUTHORS	Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R., Breese, C., Davis, A., Hopkins, J. and Freedman, R.					

TITLE		Direct Submission	
JOURNAL		Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA	
FEATURES		Location/Qualifiers	
SOURCE		1. 392	
PROMOTER		/organism="Homo sapiens"	
BASE COUNT		/db.xref="taxon:9606"	
ORIGIN		/chromosome="15"	
		/map="15q14: between D15S1031 and D15S1360"	
		/clone="CHRNA7-p"	
		1. 392	
		122 c 180 g 39 t	
Query Match		100.0%; Score 392; DB 11; Length 392;	
Best Local Similarity		100.0%; Pred. No. 2,6e-46;	
'atches 392; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
1		AGAACGCAAGGAGAGGTGAGAGCTGGCCCTTGGGAGACCCCTTGCCCTGGCCAGAGGCGG 60	
1		AGAACGCAAGGAGAGGTGAGAGCTGGCCCTTGGGAGACCCCTTGCCCTGGCCAGAGGCGG 60	
61		AGGCCGAGAGCCCGCTCGGTGAGAGCTGGGGGTGAGAGTGGCCGAGAGCTACCCAGCGCC 120	
61		AGGCCGAGAGCCCGCTCGGTGAGAGCTGGGGGTGAGAGTGGCCGAGAGCTACCCAGCGCC 120	
121		GGGAGTACCTCCCGCTCACACCTCGGGCTGCAAGTTCCCTGGGTGGCGCGCGAGAGCGTGG 180	
121		GGGAGTACCTCCCGCTCACACCTCGGGCTGCAAGTTCCCTGGGTGGCGCGCGAGAGCGTGG 180	
181		CCCCGGCTGAGAGGAGGGGGGGGGAGCGGGGGAGCGGGGGCGGGGGCTGCTACAGTGGAG 240	
181		CCCCGGCTGAGAGGAGGGGGGGGGAGCGGGGGAGCGGGGGCGGGGGCTGCTACAGTGGAG 240	
241		AGGCGGGCGGGGGCGGGGGGGCGGGGGCGCGCCCGGCTCTTAAAGCGCGCGAGCC 300	
241		AGGCGGGCGGGGGCGGGGGGGCGGGGGCGCGCCCGGCTCTTAAAGCGCGCGAGCC 300	
301		GAGCGGCGAGGTCCTCTGTGTGGCCGCGAGCGCGAGGCCCGGGCGAGACCGAGACGTGGAG 360	
301		GAGCGGCGAGGTCCTCTGTGTGGCCGCGAGCGCGAGGCCCGGGCGAGACCGAGACGTGGAG 360	
361		CGCGCGGCTCGCTGCAGCTCCGGGACTCAAC 392	
361		CGCGCGGCTCGCTGCAGCTCCGGGACTCAAC 392	
RESULT 2			
HS062436		1876 bp mRNA PRI 11-JAN-1997	
LOCUS		Human nicotinic acetylcholine receptor alpha7 subunit precursor,	
DEFINITION		mRNA, complete cds.	
ACCESSION		U62436	
NID		G1458119	
VERSION		U62436.1 GI:1458119	
KEYWORDS		human.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
REFERENCE		Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 (bases 1 to 1876)	
TITLE		Elliott,K.J., Ellis,S.B., Bercikan,K.J., Urrutia,A.,	
		Chavez-Noriega,L.E., Johnson,E.C., Vellicelid1,G. and Harpold,M.M.	
		Comparative structure of human neuronal alpha 2-alpha 7 and beta	
		2-beta 4 nicotinic acetylcholine receptor subunits and functional	
		expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and	
		beta 4 subunits	
		J. Mol. Neurosci. 7 (3), 217-228 (1996)	
JOURNAL		MEDLINE	
REFERENCE		2 (bases 1 to 1876)	
AUTHORS		Elliott,K.J.	
TITLE		Direct Submission	

JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBIA Neurosciences, Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA

FEATURES Location/Qualifiers

Source

1..1876

organism="Homo sapiens"

db_xref="taxon:9606"

/clone="KEalpha7.3"

/clone_11b="SIBIA lambda gtl1 library #3 (M. Williams)"

/cell_line="IMR32"

/cell_type="neuroblastoma"

1..72

73..138

73..1581

/codon_start=1

/product="nicotinic acetylcholine receptor alpha7 subunit

precursor="

/protein_id="AAB40114.1"

/db_xref="PID:g1458120"

/db_xref="GI:1458120"

/translation="MRCSFGGVGVALAASLLAVSLSGEFORLKYELKYKNPLERY
ANDSOPLTYEFLSLQIMDVDEKQVLTNTIMQSWTDHYLQNNVSEYGVTVRF
PDQICIKPPIILLYNSADERFATFNHNVNLSGHCQYLPGIEKSCSIDVRNPEP
VQCKIKRGSWSIGSGNSLDMQEPADISGILPNEGMDLVGIPGKSEFYCCKEPY
DVFETVYMKRRTILYGLNLILPCVILSALALVPLILPDSGKISLGTIVLSLVEF
LVAELTPIASDVPLIAQTFPACSTHRCOSLASVEMSVAPPPASNGNLIXIGRGL
LNMCAWFLLMKRPEDKRYRACONKORCSLASVEMSVAPPPASNGNLIXIGRGL
DGHVCPYDPSGVGCRMACSPTHDEHLHGQCPGEPDILKILLEVRYIANFRQC
DSEENVCSEKRFACVCDRLCMASFVETIICTIGILMSAPNFEAVSKDPA"

139..1578

/product="nicotinic acetylcholine receptor alpha7 subunit"

3'UTR 1582..1876 531 g 423 t

BASE COUNT 369 a 553 c

ORIGIN

Query Match 18.4% Score 72; DB 10; Length 1876;
Best Local Similarity 100.0%; Pred. No. 0.021; Indels 0; Gaps 0;
Matches 72; Conservative 0; Mismatches 0

QY 321 GGCGCGAGCGGAGCGCGGGGAGACAGCGAGCGAGCGCGCGCTCGTGCAGCT 380
|||||
DB 1 GGCGCGAGCGGAGCGCGGGGAGACAGCGAGCGAGCTGAGCGCGCGCTCGTGCAGCT 60
|||||

QY 381 CCGGCGACTCAAC 392
|||||

DB 61 CCGGCGACTCAAC 72

RESULT 3
AC007686
LOCUS
DEFINITION
SEQUENCE
AC007686
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

AC007686 168919 bp DNA HTG 30-MAY-1999
Homo sapiens chromosome 14 clone BAC 2289B16 map 14q24.3, LOW-PASS
SEQUENCE SAMPLING.
AC007686
94927297
AC007686.1 GI:4927297
HTG; HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 168919)
James, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,
James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,
Shaffer, T. and Hood, L.
Sequencing of human chromosome 14q24.3 region
Unpublished
2 (bases 1 to 168919)
Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,
James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,
Shaffer, T. and Hood, L.
Direct Submission
Submitted (30-MAY-1999) Multimegabase Sequencing Center, University

COMMENT

of Washington, PO BOX 357730, Seattle, WA 98195, USA
* NOTE: his record contains 192 individual.
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 965: contig of 965 bp in length
gap of unknown length
966 1866: contig of 901 bp in length
gap of unknown length
1867 2747: contig of 881 bp in length
gap of unknown length
2748 3608: contig of 861 bp in length
gap of unknown length
3609 4462: contig of 854 bp in length
gap of unknown length
4463 5345: contig of 883 bp in length
gap of unknown length
5346 6187: contig of 842 bp in length
gap of unknown length
6188 7072: contig of 885 bp in length
gap of unknown length
7073 7927: contig of 855 bp in length
gap of unknown length
7928 8797: contig of 870 bp in length
gap of unknown length
8798 9647: contig of 850 bp in length
gap of unknown length
9648 10555: contig of 908 bp in length
gap of unknown length
10556 11387: contig of 832 bp in length
gap of unknown length
11388 12285: contig of 898 bp in length
gap of unknown length
12286 13144: contig of 859 bp in length
gap of unknown length
13145 14010: contig of 866 bp in length
gap of unknown length
14011 14869: contig of 859 bp in length
gap of unknown length
14870 15756: contig of 887 bp in length
gap of unknown length
15757 16611: contig of 855 bp in length
gap of unknown length
16612 17482: contig of 871 bp in length
gap of unknown length
17483 18330: contig of 848 bp in length
gap of unknown length
18331 19206: contig of 876 bp in length
gap of unknown length
19207 20072: contig of 866 bp in length
gap of unknown length
20073 20946: contig of 874 bp in length
gap of unknown length
20947 21791: contig of 845 bp in length
gap of unknown length
21792 23944: contig of 2153 bp in length
gap of unknown length
23945 24793: contig of 849 bp in length
gap of unknown length
24794 25663: contig of 870 bp in length
gap of unknown length
25664 26504: contig of 841 bp in length
gap of unknown length
26505 27381: contig of 877 bp in length
gap of unknown length
27382 28254: contig of 873 bp in length
gap of unknown length

28255 29126: contig of 872 bp in length
gap of unknown length
29127 29963: contig of 837 bp in length
gap of unknown length
29964 30822: contig of 859 bp in length
gap of unknown length
30823 31680: contig of 856 bp in length
gap of unknown length
31681 32538: contig of 858 bp in length
gap of unknown length
32539 33390: contig of 852 bp in length
gap of unknown length
33391 34255: contig of 865 bp in length
gap of unknown length
34256 35105: contig of 850 bp in length
gap of unknown length
35106 36030: contig of 925 bp in length
gap of unknown length
36031 36869: contig of 839 bp in length
gap of unknown length
36870 37734: contig of 865 bp in length
gap of unknown length
37735 38584: contig of 850 bp in length
gap of unknown length
38585 39448: contig of 864 bp in length
gap of unknown length
39449 40283: contig of 835 bp in length
gap of unknown length
40284 41169: contig of 886 bp in length
gap of unknown length
41170 42014: contig of 845 bp in length
gap of unknown length
42015 42927: contig of 913 bp in length
gap of unknown length
42928 43788: contig of 861 bp in length
gap of unknown length
43789 44650: contig of 862 bp in length
gap of unknown length
44651 45491: contig of 841 bp in length
gap of unknown length
45492 46383: contig of 892 bp in length
gap of unknown length
46384 47232: contig of 849 bp in length
gap of unknown length
47233 48096: contig of 864 bp in length
gap of unknown length
48097 48951: contig of 855 bp in length
gap of unknown length
48952 49831: contig of 880 bp in length
gap of unknown length
49832 50676: contig of 845 bp in length
gap of unknown length
50677 51534: contig of 856 bp in length
gap of unknown length
51535 52388: contig of 854 bp in length
gap of unknown length
52389 53266: contig of 878 bp in length
gap of unknown length
53267 54108: contig of 842 bp in length
gap of unknown length
54109 55136: contig of 1028 bp in length
gap of unknown length
55137 55986: contig of 850 bp in length
gap of unknown length
55987 56874: contig of 888 bp in length
gap of unknown length
56875 57721: contig of 847 bp in length
gap of unknown length
57722 58587: contig of 866 bp in length
gap of unknown length
58588 59427: contig of 840 bp in length
gap of unknown length

* 59428 60292: contig of 865 bp in length
 * gap of unknown length
 * 60293 61149: contig of 857 bp in length
 * gap of unknown length
 * 61150 62022: contig of 873 bp in length
 * gap of unknown length
 * 62023 62858: contig of 836 bp in length
 * gap of unknown length
 * 62859 63733: contig of 875 bp in length
 * gap of unknown length
 * 63734 64579: contig of 846 bp in length
 * gap of unknown length
 * 64580 65451: contig of 872 bp in length
 * gap of unknown length
 * 65452 66289: contig of 838 bp in length
 * gap of unknown length
 * 66290 67174: contig of 885 bp in length
 * gap of unknown length
 * 67175 68020: contig of 846 bp in length
 * gap of unknown length
 * 68021 68903: contig of 883 bp in length
 * gap of unknown length
 * 68904 69757: contig of 854 bp in length
 * gap of unknown length
 * 69758 70628: contig of 871 bp in length
 * gap of unknown length
 * 70629 71466: contig of 838 bp in length
 * gap of unknown length
 * 71467 72361: contig of 895 bp in length
 * gap of unknown length
 * 72362 73206: contig of 845 bp in length

Query Match 18.3%; Score 71.8; DB 35; Length 168919;
 Best Local Similarity 51.2%; Pred. No. 0.0049;
 Matches 189; Conservative 0; Mismatches 179; Indels 1; Gaps 1;

5 CGCAGAGGAGAGGTAGACCTGCTTGCGAGCCCTGCGCTGCGCCAGAGAGCGCGAGGC 64
 DB 42401 CGNATGGGGGCGAGAGGGGCTTCGGGGGGGCTCTGCTGCTGCTGCTGCGCGGCGCGG 42460
 65 CGAGAGCCCGCTCGGTGAGAGCTGGGGGTGAGAGTCCCGGAGGCTACCCAGCGCGGGA 124
 DB 42461 GCGCGGGGTGCGGGGGGTGCGCGCGCGAGCGGGGTGCGCTGGGGCGGCTGGGCTTGGN 42520
 125 GTACTCTCCGCTCACACCTCGGGGCTGCACTTCCCTGAGTGGCGCGCGAGACGCTGCCCC 184
 DB 42521 GCGCGGCGCGCGGG 42580
 185 GCGTGAAGGAGATGCGGGCGGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 244
 DB 42581 CCGGG 42640
 245 GCGCGGG 303
 DB 42641 CCGGG 42700
 304 CCGCGAGAGTCTCTGTGGCGGCGAGCGCGCGCGCGCGCGAGACCGAGACGTGGAGCGC 363
 DB 42701 GGG 42760
 364 GCGGGGCTCG 372
 DB 42761 GCGGGGGGG 42769

RESULT 4
 LOCUS CEY39B6 342379 bp DNA HTS 09-JUN-1999
 DEFINITION Caenorhabditis elegans chromosome V clone Y39B6, WORKING DRAFT
 SEQUENCE, in unordered pieces.
 ACCESSION Z95399
 NID 95042206
 VERSION Z95399.5 GI:5042206

KEYWORDS HTG; HTGS, PHASE1.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditinae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 342379)
 AUTHORS McMurray, A.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-1999) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jesusanger.ac.uk or twenematode.wustl.edu
 COMMENT On Jun 11, 1999 this sequence version replaced gi:4725958. Order of segments is not known; 800 n's separate segments. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 SOURCE 1..342379
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="v"
 /clone="Y39B6"
 BASE COUNT 106570 a 59268 c 58805 g 109715 t 8021 others
 ORIGIN

Query Match 17.9%; Score 70; DB 34; Length 342379;
 Best Local Similarity 50.9%; Pred. No. 0.0059;
 Matches 166; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

43 GGCCTGCCAGAGCGCGCGAGAGCCGCGCTCGGTGAGACTGGGGGTGAGAGTGC 102
 DB 330422 GGG 330363
 103 CGAGAGCTTACCGAGCGCGCGAGTACCTCCGCTCACACCTCGGGCTGCACTTCCCTGGG 162
 DB 330362 GGG 330303
 163 TGCGCGCGAGACGCTGCGCCCGGCTGAGAGGATGCGGGGGGGGGGAGAGCGCGGGGGC 222
 DB 330302 AGGGGGGGGGAGGG 330243
 223 GGGGCTCTCACGTGAGAGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 282
 DB 330242 GGG 330183
 283 CTTAAGGGCGCGAGAGCGCGAGAGTCTCTGTGGCGCGAGCGCGAGGCCCGGGC 342
 DB 330182 GGG 330123
 343 GACAGCGGAGACGTGAGCGCGCGCGG 368
 DB 330122 GCGGG 330097

RESULT 5
 LOCUS HSARA7A 2087 bp mRNA PRI 01-JUN-1994
 DEFINITION H.sapiens mRNA for neuronal nicotinic acetylcholine receptor alpha-7 subunit.
 ACCESSION X70297
 NID 949606
 VERSION X70297.1 GI:496606
 KEYWORDS neuronal, nicotinic acetylcholine receptor alpha-7 subunit, human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2087)
AUTHORS Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.
TITLE Human alpha 7 acetylcholine receptor: cloning of the alpha 7 subunit from the SH-SY5Y cell line and determination of alpha 7 pharmacological properties of native receptors and functional alpha 7 homomers expressed in Xenopus oocytes
Mol. Pharmacol. 45 (3), 546-554 (1994)
JOURNAL MEDLINE 94195283
REFERENCE 2 (bases 1 to 2087)
AUTHORS Katz, M.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235 Stemmler Hall, Philadelphia, PA 19104, USA
FEATURES
source
1. 2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/clone_lib="SHSY-5Y"
/clone_lib="lambda zap II"
/clone="SHSY3"
104. 175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
104. 1612
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
/protein_id="CAA49778.1"
/db_xref="PID:9496607"
/db_xref="GI:496607"
/translation="MCSFGVWLGLASLHVSLSQGEFORLKYELVKNYNLERPV
ANDSPLTYFSLSLQIMDVDEKNQVLTNTLMQSWTDHYLQNNVSEYPKVTFRE
PDGQIMKPDILYNSADERFDATFTNTVYNSGHCQYLPPIFKSCYIDRWPFED
VOHCKLFGSNYSGMSLDLQKQEDADISGYPNGMDYLGIRGKSEPFREYECCKEEDP
DVTFTVTRRRRLYGLNLLPCVLISALALVPLPDSGKISLGLTVLSLSTVFM
LLVATIMPATSDSVPLIAQNPASTMTITGSLVTVYVYLYQHHPDGGKMKRTVRI
LLNCAWFLMKRKPGEDEKVPACQKHQRCSLAVEMSAVGPSPASNGNLLYIGRGL
DGVHCVPTPDSCVVCGRMACSEPTDHEHLHGQPEGDPDLKILEEYRYLIANFRQC
DESEAVCESEMKFACVVDRLCLMAFSVFTICTIGILMSAPNFEAVSKDFA"
mat_peptide 441 a 601 c 573 g 472 t
BASE COUNT
IGIN
Query Match 17.8%; Score 69.8; DB 10; Length 2087;
Best Local Similarity 95.9%; Pred. No. 0.041; 2; Indels 2; Gaps 2;
Matches 93; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 296 GAGCCGAGCGCGAGGTCCTCTGTGCGGAGCGAGCGCGCGGCGAGCGAGAGC 355
DB 9 GAGCCGAGCGCGAGGTCCTCTGTGCGGAGCGAGCGCGGCGAGCGAGAGC 66
QY 356 TGGAGCGCGCGGCTGCTGACGACCTCCGGGACTCAAC 392
DB 67 TGGAGCGCGCGGCTGCTGACGACCTCCGGGACTCAAC 103
RESULT 6
LOCUS HS21EA 1072 bp DNA VRL 15-MAR-1990
DEFINITION Herpes simplex virus type 2 immediate-early (IE5) protein mRNA, 5'
end.
ACCESSION M29385
NID 9330283
VERSION M29385.1 GI:330283
KEYWORDS

SOURCE
ORGANISM Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed
in baby hamster kidney clone 21 (C13) cells.
REFERENCE 1 (bases 1 to 1072)
AUTHORS Whitton, J. L. and Clements, J. B.
TITLE The junctions between the repetitive and the short unique sequences
of the herpes simplex virus genome are determined by the
polypeptide-coding regions of two spliced immediate-early mRNAs
J. Gen. Virol. 65, 451-466 (1984)
JOURNAL MEDLINE 84137573
FEATURES
source
1. 1072
/organism="human herpesvirus 2"
/db_xref="taxon:10310"
prim_transcript 95. >1072
/note="IE-5 mRNA and intron"
344. 866
/note="IE-5 intron"
897. >1072
/note="immediate early protein 5"
CDS
/codon_start=1
/protein_id="AAA45849.1"
/db_xref="PID:9330284"
/db_xref="GI:330284"
/translation="MSNALKRTTDMPSDSSRCHRTHTYGVCAEIKHREDEARNTAV
TDPEPLICPDVPR"
BASE COUNT 132 a 322 c 543 g 75 t
ORIGIN
Query Match 17.3%; Score 68; DB 17; Length 1072;
Best Local Similarity 50.6%; Pred. No. 0.091;
Matches 164; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 37 GCCCTGCGCTCGCCAGAGGCGCGAGAGCCCGCTCGTGAGACTGGGGGTGA 96
DB 366 GCCGGGGGAGCGGGCGGGGAGCGGGCGGGGAGCGGGCGGGGAGCGGGCGGGG 425
QY 97 GGTGCGCGGAGCGTACCGAGCGCGCGGAGTACCTCCGCTCACCTCGGCTGAGTTC 156
DB 426 GACGGGCGCGGGGAGCGGGCGCGGGGAGCGGGCGGGGAGCGGGCGGGGAGCGGGG 485
QY 157 OCTGGGTGGCGCGGAGAGCGTGGCGCGCGCGCGCGCGAGAGTGGCGGGGAGCGGGGCGC 216
DB 486 GCGGGGGGCGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 545
QY 217 GGGGCGGGGCTCTGACAGTGGAGAGCGCGCGGGGCGGGGCGGGGCGGGCGCGCC 276
DB 546 CGGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGCGC 605
QY 277 CGGCTCTTAAAGCGCGCGAGCGCGAGCGCGAGTGGCTCTGTGCGCGGAGCGCGAGCGC 336
DB 606 GGGGAGCGGGGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 665
QY 337 CCGGGGAGCGCGGAGCGAGTGGAG 360
DB 666 CCGGGGAGCGGGGAGCGGGGAGCGGGG 689
RESULT 7
LOCUS HS21E 1560 bp DNA VRL 15-MAR-1990
DEFINITION Herpes simplex virus type 2 immediate-early (IE4) protein mRNA, 5'
end.
ACCESSION M29384
NID 9330281
VERSION M29384.1 GI:330281
KEYWORDS
SOURCE Herpes simplex virus type 2 (strain HGS Timbury, 1971) DNA, passed
in baby hamster kidney clone 21 (C13) cells.
ORGANISM human herpesvirus 2


```

repeat_region /note="AluY repeat: matches 13. .294 of consensus"
10382..10428
/note="AluX/g repeat: matches 245. .291 of consensus;
incomplete repeat"
repeat_region /note="L1MB8 repeat: matches 920. .859 of consensus"
10652..10713
/note="L1MB8 repeat: matches 1. .255 of consensus;
incomplete repeat"
repeat_region /note="AluX repeat: matches 1. .145 of consensus;
incomplete repeat"
11022..11168
/note="AluX repeat: matches 1. .145 of consensus;
incomplete repeat"
11169..11463
/note="AluX repeat: matches 1. .296 of consensus"
11466..11635
/note="AluX repeat: matches 133. .302 of consensus;
incomplete repeat"
11730..12317
/note="L1 repeat: matches 4729. .5317 of consensus"
12321..12616
/note="AluY repeat: matches 1. .296 of consensus"
12633..12761
/note="AluX repeat: matches 1. .131 of consensus;
incomplete repeat"
12779..13073
/note="AluX repeat: matches 1. .297 of consensus"
join(13773. .13867,14755. .14959,15135. .15263)
/codon_start=1
/product="zeta-globin 1"
/protein_id="CAB06552.1"
/db_xref="PID:g300370"
/db_xref="PID:g1817576"
/db_xref="GI:1817576"
/db_xref="SWISS-PROT:P02008"
/translation="MSLTERTITVSMMAKISQADPTGTETLEFLSHPOQTYE
PHFDLPGSAQARAGSKYVAAGVAVASIDIGALSKLSLHAYILRVDPVNFKL
SHCLVTLARFPADFTAEAHAMKRFSLVSSVLETKR"
13923..14090
/note="L12 copies of 14 mer 89 & conserved"
14618..15730
/note="Putative Cpg Island"
14964..15071
/note="3 copies of 36 mer 83 & conserved"
15747..15875
/note="MIR2 repeat: matches 145. .10 of consensus"
16241..16541
/note="AluY repeat: matches 301. .1 of consensus"
16892..17027
/note="AluY repeat: matches 1. .136 of consensus;
incomplete repeat"
17029..17326
/note="AluY repeat: matches 1. .297 of consensus"
17368..17506
/note="AluY repeat: matches 10. .151 of consensus;
incomplete repeat"
17945..18086
/note="AluY repeat: matches 153. .294 of consensus;
incomplete repeat"
18106..18426
/note="AluY repeat: matches 1. .301 of consensus"
18715..18833
/note="L1MB7 repeat: matches 744. .865 of consensus"
18834..19130
/note="AluX repeat: matches 1. .296 of consensus"
19186..19449
/note="MER42C repeat: matches 1506. .1230 of consensus"
19431..19593
/note="HUMAR1 repeat: matches 1221. .1058 of consensus"
19601..19838
/note="AluX repeat: matches 300. .63 of consensus;
incomplete repeat"
19840..19983
/note="AluX repeat: matches 154. .300 of consensus;

```

```

incomplete repeat"
repeat_region /note="4 copies of 37 mer 93 & conserved"
20709..20856
repeat_region /note="2 copies of 36 mer 99 & conserved"
20881..20952
repeat_region /note="3 copies of 37 mer 97 & conserved"
20919..21029
repeat_region /note="2 copies of 36 mer 99 & conserved"
20966..21067
repeat_region /note="4 copies of 37 mer 95 & conserved"
21039..21186
repeat_region /note="3 copies of 36 mer 97 & conserved"
21210..21323
repeat_region /note="2 copies of 36 mer 97 & conserved"
21325..21396
repeat_region /note="2 copies of 36 mer 97 & conserved"
21360..21433
repeat_region /note="2 copies of 37 mer 97 & conserved"
21396..21469
repeat_region /note="2 copies of 36 mer 97 & conserved"
21471..21542
repeat_region /note="2 copies of 36 mer 96 & conserved"
21546..21617
repeat_region /note="2 copies of 36 mer 96 & conserved"
21579..21726
repeat_region /note="4 copies of 37 mer 94 & conserved"
22591..22661
/note="AluX repeat: matches 302. .232 of consensus;
incomplete repeat"
22766..23070
/note="AluX repeat: matches 1. .302 of consensus"
join(23985. .24079,25345. .25549,25891. .26019)
/codon_start=1
/pseudo
/product="zeta-globin 2"
/db_xref="PID:e300371"
24135..24680
/note="39 copies of 14 mer 90 & conserved"
25208..27854
/note="Putative Cpg Island"
p1rn_transcript <26871..>27290
/note="match: multiple ESTs; similar to HEMOGLOBIN
Alpha-D; match: 5' EST H57023 clone 204625; match: 5' EST
R92558 clone 196372; match: 3' EST H67154 clone 211156"

```

Query Match 16.1%; Score 63.2; DB 10; Length 43058;
 Best Local Similarity 52.7%; Pred. NO. 0.12; Indels 0; Gaps 0;
 Matches 137; Conservative 0; Mismatches 123;

```

QY 24 CTGGCTTGAGGAGCCCTGAGCCAGAGCGGAGAGCCCGCTCGGTGA 83
DB 25589 CGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGG 25648
QY 84 GACTGGGGGTGAGTCCCGGAGCGTACCAAGCCCGGAGATCTCCGCTCACT 143
DB 25649 GGGCGGGGTGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGG 25708
QY 144 CGGGCTGACGTCCCTGAGTGGGCGCGAGACGCTGGGCGGCTGAGAGCATGCGGGG 203
DB 25709 GGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 25768
QY 204 CGGGAGCGGGGCGGGGCGGGGCGGTCTGACATGAGAGAGCGGCGGGGCGGGG 263
DB 25769 CGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 25828
QY 264 GGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 283
DB 25829 GGGGTGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 25848

```

RESULT 9
 HUMBAR3 HUMBAR3 2665 bp DNA PRI 18-SEP-1997
 LOCUS
 DEFINITION Human alpha globin gene cluster on chromosome 16: psi2eta

pseudogene.
 ACCESSION J00184
 NID 9183792
 VERSION J00184.1 GI:183792
 KEYWORDS gene duplication; pseudogene; repeat region; zeta-globin.
 SEGMENT 3 of 4
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 2685)
 PROUDFOOT,N.J., GILL,A. and MANIATIS,T.
 The structure of the human zeta-globin gene and a closely linked,
 nearly identical pseudogene
 Cell 31 (3 Pt 2), 553-563 (1982)
 JOURNAL 83129370
 MEDLINE
 COMMENT The human alpha globin gene cluster, located on the short arm of chromosome 16, spans about 30 kb and includes the following five loci:
 5'-zeta-pseudozeta-pseudozeta-alpha-1-alpha-2-alpha-1-3'. The gene sequence shown below is considered a pseudogene because of the terminating codon at base 429. There are only two additional single-base differences between the coding sequences of the zeta and pseudozeta genes, although there are conspicuous differences in the 5' and 3' flanking regions. Moreover the first intron of this locus demonstrates length variation, apparently stemming from the repeat sequence 'acagtgagagag' which is also found in the zeta gene and, in a similar form, in the 5' flank of the human insulin gene. The repeat sequence 'cgagag' is characteristic of the second introns of the zeta genes [1]. Typical promoter elements 'ccat' and 'tata' are found at bases 291 and 326.
 FEATURES
 Source Location/Qualifiers
 1. 2685
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="16p13.3"
 356..505
 /gene="phbz"
 /number=1
 join(411..505,1767..1971,2313..2441)
 /gene="phbz"
 /pseudo
 /codon_start=1
 join(411..505,1767..1971,2313..2441)
 /gene="phbz"
 /number=1
 /note="repeat sequence copy 1"
 1767..1971
 /gene="phbz"
 /number=2
 1972..2312
 /gene="phbz"
 /number=2
 1977..1981
 /note="repeat sequence, copy 1"
 2313..2541
 /gene="phbz"
 /number=3
 2521..2526
 /note="phbz mRNA polyadenylation signal"
 BASE COUNT 454 a 720 c 1155 g 356 t
 ORIGIN About 1 kb after segment 2.
 Query Match 16.1%; Score 63.2; DB 10; Length 2685;
 Best Local Similarity 52.7%; Pred. No. 0.31;

Matches 137; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
 QY 24 CTGGCTTGGCAACCCCTGCGCCAGAGCGCGAGAGCCCGCGTGGTGA 83
 Db 2011 CGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 2070
 QY 84 GACHTGGGGTGGAGTGCCTCCGAGACCGCGGAGTAGTACTCCGCTCACACT 143
 Db 2071 GGGCGGGGTCCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 2130
 QY 144 CGGCTCAGTTCCTGGGTGGCCGCGAGACGCTGGCCCGGCTGAGAGATCGGGG 203
 Db 2131 GGGCGGGGGCGGGCGGGGAGGGCTGGGCGGGCGGGCGGGCGGGCGGGCG 2190
 QY 204 CGGGAGCGGGGGCGGGCGGGCGGCTCTGCTACGTGAGAGCGCGCGGGCGGGCG 263
 Db 2191 CCGGGCGGGGTCTCGGGCGGGGTGCGGGCGGGCGGGCGGGCGGGCGGGGT 2250
 QY 264 GGGGGCGGGCGGGCGGGCGGCTCC 283
 Db 2251 GGGGTCTGGGGCGGGCGGGCC 2270
 RESULT 10
 LOCUS HSV2HG52 DNA VRL 04-DEC-1998
 DEFINITION Herpes simplex virus type 2 (strain HG52), complete genome.
 ACCESSION 286099
 NID 91869820
 VERSION 286099.1 GI:1869820
 KEYWORDS capsid associated tegument protein; capsid protein; complete genome; deoxyribonuclease; deoxyribonucleic acid triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capsid protein; membrane-associated phosphoprotein; minor capsid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; R1 gene; R2 gene; R3 gene; tegument protein; thymidine kinase; U1 gene; U10 gene; U11 gene; U12 gene; U13 gene; U14 gene; U15 gene; U16 gene; U17 gene; U18 gene; U19 gene; U20 gene; U21 gene; U22 gene; U23 gene; U24 gene; U25 gene; U26 gene; U26.5 gene; U27 gene; U28 gene; U29 gene; U30 gene; U31 gene; U32 gene; U33 gene; U34 gene; U35 gene; U36 gene; U37 gene; U38 gene; U39 gene; U40 gene; U41 gene; U42 gene; U43 gene; U44 gene; U45 gene; U46 gene; U47 gene; U48 gene; U49 gene; U50 gene; U51 gene; U52 gene; U53 gene; U54 gene; U55 gene; U56 gene; U57 gene; U58 gene; U59 gene; U60 gene; U61 gene; U62 gene; U63 gene; U64 gene; U65 gene; U66 gene; U67 gene; U68 gene; U69 gene; U70 gene; U71 gene; U72 gene; U73 gene; U74 gene; U75 gene; U76 gene; U77 gene; U78 gene; U79 gene; U80 gene; U81 gene; U82 gene; U83 gene; U84 gene; U85 gene; U86 gene; U87 gene; U88 gene; U89 gene; U90 gene; U91 gene; U92 gene; U93 gene; U94 gene; U95 gene; U96 gene; U97 gene; U98 gene; U99 gene; U100 gene; U101 gene; U102 gene; U103 gene; U104 gene; U105 gene; U106 gene; U107 gene; U108 gene; U109 gene; U110 gene; U111 gene; U112 gene; U113 gene; U114 gene; U115 gene; U116 gene; U117 gene; U118 gene; U119 gene; U120 gene; U121 gene; U122 gene; U123 gene; U124 gene; U125 gene; U126 gene; U126.5 gene; U127 gene; U128 gene; U129 gene; U130 gene; U131 gene; U132 gene; U133 gene; U134 gene; U135 gene; U136 gene; U137 gene; U138 gene; U139 gene; U140 gene; U141 gene; U142 gene; U143 gene; U144 gene; U145 gene; U146 gene; U147 gene; U148 gene; U149 gene; U150 gene; U151 gene; U152 gene; U153 gene; U154 gene; U155 gene; U156 gene; U157 gene; U158 gene; U159 gene; U160 gene; U161 gene; U162 gene; U163 gene; U164 gene; U165 gene; U166 gene; U167 gene; U168 gene; U169 gene; U170 gene; U171 gene; U172 gene; U173 gene; U174 gene; U175 gene; U176 gene; U177 gene; U178 gene; U179 gene; U180 gene; U181 gene; U182 gene; U183 gene; U184 gene; U185 gene; U186 gene; U187 gene; U188 gene; U189 gene; U190 gene; U191 gene; U192 gene; U193 gene; U194 gene; U195 gene; U196 gene; U197 gene; U198 gene; U199 gene; U200 gene; U201 gene; U202 gene; U203 gene; U204 gene; U205 gene; U206 gene; U207 gene; U208 gene; U209 gene; U210 gene; U211 gene; U212 gene; U213 gene; U214 gene; U215 gene; U216 gene; U217 gene; U218 gene; U219 gene; U220 gene; U221 gene; U222 gene; U223 gene; U224 gene; U225 gene; U226 gene; U226.5 gene; U227 gene; U228 gene; U229 gene; U230 gene; U231 gene; U232 gene; U233 gene; U234 gene; U235 gene; U236 gene; U237 gene; U238 gene; U239 gene; U240 gene; U241 gene; U242 gene; U243 gene; U244 gene; U245 gene; U246 gene; U247 gene; U248 gene; U249 gene; U250 gene; U251 gene; U252 gene; U253 gene; U254 gene; U255 gene; U256 gene; U257 gene; U258 gene; U259 gene; U260 gene; U261 gene; U262 gene; U263 gene; U264 gene; U265 gene; U266 gene; U267 gene; U268 gene; U269 gene; U270 gene; U271 gene; U272 gene; U273 gene; U274 gene; U275 gene; U276 gene; U277 gene; U278 gene; U279 gene; U280 gene; U281 gene; U282 gene; U283 gene; U284 gene; U285 gene; U286 gene; U287 gene; U288 gene; U289 gene; U290 gene; U291 gene; U292 gene; U293 gene; U294 gene; U295 gene; U296 gene; U297 gene; U298 gene; U299 gene; U300 gene; U301 gene; U302 gene; U303 gene; U304 gene; U305 gene; U306 gene; U307 gene; U308 gene; U309 gene; U310 gene; U311 gene; U312 gene; U313 gene; U314 gene; U315 gene; U316 gene; U317 gene; U318 gene; U319 gene; U320 gene; U321 gene; U322 gene; U323 gene; U324 gene; U325 gene; U326 gene; U326.5 gene; U327 gene; U328 gene; U329 gene; U330 gene; U331 gene; U332 gene; U333 gene; U334 gene; U335 gene; U336 gene; U337 gene; U338 gene; U339 gene; U340 gene; U341 gene; U342 gene; U343 gene; U344 gene; U345 gene; U346 gene; U347 gene; U348 gene; U349 gene; U350 gene; U351 gene; U352 gene; U353 gene; U354 gene; U355 gene; U356 gene; U357 gene; U358 gene; U359 gene; U360 gene; U361 gene; U362 gene; U363 gene; U364 gene; U365 gene; U366 gene; U367 gene; U368 gene; U369 gene; U370 gene; U371 gene; U372 gene; U373 gene; U374 gene; U375 gene; U376 gene; U377 gene; U378 gene; U379 gene; U380 gene; U381 gene; U382 gene; U383 gene; U384 gene; U385 gene; U386 gene; U387 gene; U388 gene; U389 gene; U390 gene; U391 gene; U392 gene; U393 gene; U394 gene; U395 gene; U396 gene; U397 gene; U398 gene; U399 gene; U400 gene; U401 gene; U402 gene; U403 gene; U404 gene; U405 gene; U406 gene; U407 gene; U408 gene; U409 gene; U410 gene; U411 gene; U412 gene; U413 gene; U414 gene; U415 gene; U416 gene; U417 gene; U418 gene; U419 gene; U420 gene; U421 gene; U422 gene; U423 gene; U424 gene; U425 gene; U426 gene; U426.5 gene; U427 gene; U428 gene; U429 gene; U430 gene; U431 gene; U432 gene; U433 gene; U434 gene; U435 gene; U436 gene; U437 gene; U438 gene; U439 gene; U440 gene; U441 gene; U442 gene; U443 gene; U444 gene; U445 gene; U446 gene; U447 gene; U448 gene; U449 gene; U450 gene; U451 gene; U452 gene; U453 gene; U454 gene; U455 gene; U456 gene; U457 gene; U458 gene; U459 gene; U460 gene; U461 gene; U462 gene; U463 gene; U464 gene; U465 gene; U466 gene; U467 gene; U468 gene; U469 gene; U470 gene; U471 gene; U472 gene; U473 gene; U474 gene; U475 gene; U476 gene; U477 gene; U478 gene; U479 gene; U480 gene; U481 gene; U482 gene; U483 gene; U484 gene; U485 gene; U486 gene; U487 gene; U488 gene; U489 gene; U490 gene; U491 gene; U492 gene; U493 gene; U494 gene; U495 gene; U496 gene; U497 gene; U498 gene; U499 gene; U500 gene; U501 gene; U502 gene; U503 gene; U504 gene; U505 gene; U506 gene; U507 gene; U508 gene; U509 gene; U510 gene; U511 gene; U512 gene; U513 gene; U514 gene; U515 gene; U516 gene; U517 gene; U518 gene; U519 gene; U520 gene; U521 gene; U522 gene; U523 gene; U524 gene; U525 gene; U526 gene; U526.5 gene; U527 gene; U528 gene; U529 gene; U530 gene; U531 gene; U532 gene; U533 gene; U534 gene; U535 gene; U536 gene; U537 gene; U538 gene; U539 gene; U540 gene; U541 gene; U542 gene; U543 gene; U544 gene; U545 gene; U546 gene; U547 gene; U548 gene; U549 gene; U550 gene; U551 gene; U552 gene; U553 gene; U554 gene; U555 gene; U556 gene; U557 gene; U558 gene; U559 gene; U560 gene; U561 gene; U562 gene; U563 gene; U564 gene; U565 gene; U566 gene; U567 gene; U568 gene; U569 gene; U570 gene; U571 gene; U572 gene; U573 gene; U574 gene; U575 gene; U576 gene; U577 gene; U578 gene; U579 gene; U580 gene; U581 gene; U582 gene; U583 gene; U584 gene; U585 gene; U586 gene; U587 gene; U588 gene; U589 gene; U590 gene; U591 gene; U592 gene; U593 gene; U594 gene; U595 gene; U596 gene; U597 gene; U598 gene; U599 gene; U600 gene; U601 gene; U602 gene; U603 gene; U604 gene; U605 gene; U606 gene; U607 gene; U608 gene; U609 gene; U610 gene; U611 gene; U612 gene; U613 gene; U614 gene; U615 gene; U616 gene; U617 gene; U618 gene; U619 gene; U620 gene; U621 gene; U622 gene; U623 gene; U624 gene; U625 gene; U626 gene; U626.5 gene; U627 gene; U628 gene; U629 gene; U630 gene; U631 gene; U632 gene; U633 gene; U634 gene; U635 gene; U636 gene; U637 gene; U638 gene; U639 gene; U640 gene; U641 gene; U642 gene; U643 gene; U644 gene; U645 gene; U646 gene; U647 gene; U648 gene; U649 gene; U650 gene; U651 gene; U652 gene; U653 gene; U654 gene; U655 gene; U656 gene; U657 gene; U658 gene; U659 gene; U660 gene; U661 gene; U662 gene; U663 gene; U664 gene; U665 gene; U666 gene; U667 gene; U668 gene; U669 gene; U670 gene; U671 gene; U672 gene; U673 gene; U674 gene; U675 gene; U676 gene; U677 gene; U678 gene; U679 gene; U680 gene; U681 gene; U682 gene; U683 gene; U684 gene; U685 gene; U686 gene; U687 gene; U688 gene; U689 gene; U690 gene; U691 gene; U692 gene; U693 gene; U694 gene; U695 gene; U696 gene; U697 gene; U698 gene; U699 gene; U700 gene; U701 gene; U702 gene; U703 gene; U704 gene; U705 gene; U706 gene; U707 gene; U708 gene; U709 gene; U710 gene; U711 gene; U712 gene; U713 gene; U714 gene; U715 gene; U716 gene; U717 gene; U718 gene; U719 gene; U720 gene; U721 gene; U722 gene; U723 gene; U724 gene; U725 gene; U726 gene; U726.5 gene; U727 gene; U728 gene; U729 gene; U730 gene; U731 gene; U732 gene; U733 gene; U734 gene; U735 gene; U736 gene; U737 gene; U738 gene; U739 gene; U740 gene; U741 gene; U742 gene; U743 gene; U744 gene; U745 gene; U746 gene; U747 gene; U748 gene; U749 gene; U750 gene; U751 gene; U752 gene; U753 gene; U754 gene; U755 gene; U756 gene; U757 gene; U758 gene; U759 gene; U760 gene; U761 gene; U762 gene; U763 gene; U764 gene; U765 gene; U766 gene; U767 gene; U768 gene; U769 gene; U770 gene; U771 gene; U772 gene; U773 gene; U774 gene; U775 gene; U776 gene; U777 gene; U778 gene; U779 gene; U780 gene; U781 gene; U782 gene; U783 gene; U784 gene; U785 gene; U786 gene; U787 gene; U788 gene; U789 gene; U790 gene; U791 gene; U792 gene; U793 gene; U794 gene; U795 gene; U796 gene; U797 gene; U798 gene; U799 gene; U800 gene; U801 gene; U802 gene; U803 gene; U804 gene; U805 gene; U806 gene; U807 gene; U808 gene; U809 gene; U810 gene; U811 gene; U812 gene; U813 gene; U814 gene; U815 gene; U816 gene; U817 gene; U818 gene; U819 gene; U820 gene; U821 gene; U822 gene; U823 gene; U824 gene; U825 gene; U826 gene; U826.5 gene; U827 gene; U828 gene; U829 gene; U830 gene; U831 gene; U832 gene; U833 gene; U834 gene; U835 gene; U836 gene; U837 gene; U838 gene; U839 gene; U840 gene; U841 gene; U842 gene; U843 gene; U844 gene; U845 gene; U846 gene; U847 gene; U848 gene; U849 gene; U850 gene; U851 gene; U852 gene; U853 gene; U854 gene; U855 gene; U856 gene; U857 gene; U858 gene; U859 gene; U860 gene; U861 gene; U862 gene; U863 gene; U864 gene; U865 gene; U866 gene; U867 gene; U868 gene; U869 gene; U870 gene; U871 gene; U872 gene; U873 gene; U874 gene; U875 gene; U876 gene; U877 gene; U878 gene; U879 gene; U880 gene; U881 gene; U882 gene; U883 gene; U884 gene; U885 gene; U886 gene; U887 gene; U888 gene; U889 gene; U890 gene; U891 gene; U892 gene; U893 gene; U894 gene; U895 gene; U896 gene; U897 gene; U898 gene; U899 gene; U900 gene; U901 gene; U902 gene; U903 gene; U904 gene; U905 gene; U906 gene; U907 gene; U908 gene; U909 gene; U910 gene; U911 gene; U912 gene; U913 gene; U914 gene; U915 gene; U916 gene; U917 gene; U918 gene; U919 gene; U920 gene; U921 gene; U922 gene; U923 gene; U924 gene; U925 gene; U926 gene; U926.5 gene; U927 gene; U928 gene; U929 gene; U930 gene; U931 gene; U932 gene; U933 gene; U934 gene; U935 gene; U936 gene; U937 gene; U938 gene; U939 gene; U940 gene; U941 gene; U942 gene; U943 gene; U944 gene; U945 gene; U946 gene; U947 gene; U948 gene; U949 gene; U950 gene; U951 gene; U952 gene; U953 gene; U954 gene; U955 gene; U956 gene; U957 gene; U958 gene; U959 gene; U960 gene; U961 gene; U962 gene; U963 gene; U964 gene; U965 gene; U966 gene; U967 gene; U968 gene; U969 gene; U970 gene; U971 gene; U972 gene; U973 gene; U974 gene; U975 gene; U976 gene; U977 gene; U978 gene; U979 gene; U980 gene; U981 gene; U982 gene; U983 gene; U984 gene; U985 gene; U986 gene; U987 gene; U988 gene; U989 gene; U990 gene; U991 gene; U992 gene; U993 gene; U994 gene; U995 gene; U996 gene; U997 gene; U998 gene; U999 gene; U1000 gene; U1001 gene; U1002 gene; U1003 gene; U1004 gene; U1005 gene; U1006 gene; U1007 gene; U1008 gene; U1009 gene; U1010 gene; U1011 gene; U1012 gene; U1013 gene; U1014 gene; U1015 gene; U1016 gene; U1017 gene; U1018 gene; U1019 gene; U1020 gene; U1021 gene; U1022 gene; U1023 gene; U1024 gene; U1025 gene; U1026 gene; U1026.5 gene; U1027 gene; U1028 gene; U1029 gene; U1030 gene; U1031 gene; U1032 gene; U1033 gene; U1034 gene; U1035 gene; U1036 gene; U1037 gene; U1038 gene; U1039 gene; U1040 gene; U1041 gene; U1042 gene; U1043 gene; U1044 gene; U1045 gene; U1046 gene; U1047 gene; U1048 gene; U1049 gene; U1050 gene; U1051 gene; U1052 gene; U1053 gene; U1054 gene; U1055 gene; U1056 gene; U1057 gene; U1058 gene; U1059 gene; U1060 gene; U1061 gene; U1062 gene; U1063 gene; U1064 gene; U1065 gene; U1066 gene; U1067 gene; U1068 gene; U1069 gene; U1070 gene; U1071 gene; U1072 gene; U1073 gene; U1074 gene; U1075 gene; U1076 gene; U1077 gene; U1078 gene; U1079 gene; U1080 gene; U1081 gene; U1082 gene; U1083 gene; U1084 gene; U1085 gene; U1086 gene; U1087 gene; U1088 gene; U1089 gene; U1090 gene; U1091 gene; U1092 gene; U1093 gene; U1094 gene; U1095 gene; U1096 gene; U1097 gene; U1098 gene; U1099 gene; U1100 gene; U1101 gene; U1102 gene; U1103 gene; U1104 gene; U1105 gene; U1106 gene; U1107 gene; U1108 gene; U1109 gene; U1110 gene; U1111 gene; U1112 gene; U1113 gene; U1114 gene; U1115 gene; U1116 gene; U1117 gene; U1118 gene; U1119 gene; U1120 gene; U1121 gene; U1122 gene; U1123 gene; U1124 gene; U1125 gene; U1126 gene; U1126.5 gene; U1127 gene; U1128 gene; U1129 gene; U1130 gene; U1131 gene; U1132 gene; U1133 gene; U1134 gene; U1135 gene; U1136 gene; U1137 gene; U1138 gene; U1139 gene; U1140 gene; U1141 gene; U1142 gene; U1143 gene; U1144 gene; U1145 gene; U1146 gene; U1147 gene; U1148 gene; U1149 gene; U1150 gene; U1151 gene; U1152 gene; U1153 gene; U1154 gene; U1155 gene; U1156 gene; U1157 gene; U1158 gene; U1159 gene; U1160 gene; U1161 gene; U1162 gene; U1163 gene; U1164 gene; U1165 gene; U1166 gene; U1167 gene; U1168 gene; U1169 gene; U1170 gene; U1171 gene; U1172 gene; U1173 gene; U1174 gene; U1175 gene; U1176 gene; U1177 gene; U1178 gene; U1179 gene; U1180 gene; U1181 gene; U1182 gene; U1183 gene; U1184 gene; U1185 gene; U1186 gene; U1187 gene; U1188 gene; U1189 gene; U1190 gene; U1191 gene; U1192 gene; U1193 gene; U1194 gene; U1195 gene; U1196 gene; U1197 gene; U1198 gene; U1199 gene; U1200 gene; U1201 gene; U1202 gene; U1203 gene; U1204 gene; U1205 gene; U1206 gene; U1206.5 gene; U1207 gene; U1208 gene; U1209 gene; U1210 gene; U1211 gene; U1212 gene; U1213 gene; U1214 gene; U1215 gene; U1216 gene; U1217 gene; U1218 gene; U1219 gene; U1220 gene; U1221 gene; U1222 gene; U1223 gene; U1224 gene; U1225 gene; U1226 gene; U1226.5 gene; U1227 gene; U1228 gene; U1229 gene; U1230 gene; U1231 gene; U1232 gene; U1233 gene; U1234 gene; U1235 gene; U1236 gene; U1236.5 gene; U1237 gene; U1238 gene; U1239 gene; U1240 gene; U1241 gene; U1242 gene; U1243 gene; U1244 gene; U1245 gene; U1246 gene; U1247 gene; U1248 gene; U1249 gene; U1250 gene; U1251 gene; U1252 gene; U1253 gene; U1254 gene; U1255 gene; U1256 gene; U1257 gene; U1258 gene; U1259 gene; U1260 gene; U1261 gene; U1262 gene; U1263 gene; U1264 gene; U1265 gene; U1266 gene; U1267 gene; U1268 gene; U1269 gene; U1270 gene; U1271 gene; U1272 gene; U1273 gene; U1274 gene; U1275 gene; U1276 gene; U1277 gene; U1278 gene; U1279 gene; U1280 gene; U1281 gene; U1282 gene; U1283 gene; U1284 gene; U1285 gene; U1286 gene; U1287 gene; U1288 gene; U1289 gene; U1290 gene; U1291 gene; U1292 gene; U1293 gene; U1294 gene; U1295 gene; U1296 gene; U1297 gene; U1298 gene; U1299 gene; U1300 gene; U1301 gene; U1302 gene; U1303 gene; U1304 gene; U1305 gene; U1306 gene; U1307 gene; U1308 gene; U1309 gene; U1310 gene; U1311 gene; U1312 gene; U1313 gene; U1314 gene; U1315 gene; U1316 gene; U1317 gene; U1318 gene; U1319 gene; U1320 gene; U1321 gene; U1322 gene; U1323 gene; U1324 gene; U1325 gene; U1326 gene; U1326.5 gene; U1327 gene; U1328 gene; U1329 gene; U1330 gene; U1331 gene; U1332 gene; U1333 gene; U1334 gene; U1335 gene; U1336 gene; U1337 gene; U1338 gene; U1339 gene; U1340 gene; U1341 gene; U1342 gene; U1343 gene; U1344 gene; U1345 gene; U1346 gene; U1347 gene; U1348 gene; U1349 gene; U1350 gene; U1351 gene; U1352 gene; U1353 gene; U1354 gene; U1355 gene; U1356 gene; U1357 gene; U1358 gene; U1359 gene; U1360 gene; U1361 gene; U1362 gene; U1363 gene; U1364 gene; U1365 gene; U1366 gene; U1367 gene; U1368 gene; U1369 gene; U1370 gene; U1371 gene; U1372 gene; U1373 gene; U1374 gene; U1375 gene; U1376 gene; U1377 gene; U1378 gene; U1379 gene; U1380 gene; U1381 gene; U1382 gene; U1383 gene; U1384 gene; U1385 gene; U1386 gene; U1387 gene; U1388 gene; U1389 gene; U1390 gene; U1391 gene; U1392 gene; U1393 gene; U1394 gene; U1395 gene; U1396 gene; U1397 gene; U1398 gene; U1399 gene; U1400 gene; U1401 gene; U1402 gene; U1403 gene; U1404 gene; U1405 gene; U1406 gene; U1407 gene; U1408 gene; U1409 gene; U1410 gene; U1411 gene; U1412 gene; U1413 gene; U1414 gene; U1415 gene; U1416 gene; U1417 gene; U1418 gene; U1419 gene; U1420 gene; U1421 gene; U1422 gene; U1423 gene; U1424 gene; U1425 gene; U1426 gene; U1426.5 gene; U1427 gene; U1428 gene; U1429 gene; U1430 gene; U1431 gene; U1432 gene; U1433 gene; U1434 gene; U1435 gene; U1436 gene; U1437 gene; U1438 gene; U1439 gene; U1440 gene; U1441 gene; U1442 gene; U1443 gene; U1444 gene; U1445 gene; U1446 gene; U1447 gene; U1448 gene; U1449 gene; U1450 gene; U1451 gene; U1452 gene; U1453 gene; U1454 gene; U1455 gene; U1456 gene; U1457 gene; U1458 gene; U1459 gene; U1460 gene; U1461 gene; U1462 gene; U1463 gene; U1464 gene; U1465 gene; U1466 gene; U1467 gene; U1468 gene; U1469 gene; U1470 gene; U1471 gene; U1472 gene; U1473 gene; U1474 gene; U1475 gene; U1476 gene; U1477 gene; U1478 gene; U1479 gene; U1480 gene; U1481 gene; U1482 gene; U1483 gene; U1484 gene; U1485 gene; U1486 gene; U1487 gene; U1488 gene; U1489 gene; U1490 gene; U1491 gene; U1492 gene; U1493 gene; U1494 gene; U1495 gene; U1496 gene; U1497 gene; U1498 gene; U1499 gene; U1500 gene; U1501 gene; U1502 gene; U1503 gene; U1504 gene; U1505 gene; U1506 gene; U1507 gene; U1508 gene; U1509 gene; U1510 gene; U1511 gene; U1512 gene; U1513 gene; U1514 gene; U1515 gene; U1516 gene; U1517 gene; U1518 gene; U1519 gene; U1520 gene; U1521 gene; U1522 gene; U1523 gene; U1524 gene; U1525 gene; U1526 gene; U1526.5 gene; U1527 gene; U1528 gene; U1529 gene; U1530 gene; U1531 gene; U1532 gene; U1533 gene; U1534 gene; U1535 gene; U1536 gene; U1537 gene; U1538 gene; U1539 gene; U1540 gene; U1541 gene; U1542 gene; U1543 gene; U1544 gene; U1545 gene; U1546 gene; U1547 gene; U1548 gene; U1549 gene; U1550 gene; U1551 gene; U1552 gene; U1553 gene; U1554 gene; U1555 gene; U1556 gene; U1557 gene; U1558 gene; U1559 gene; U1560 gene; U1561 gene; U1562 gene; U1563 gene; U1564 gene; U1565 gene; U1566 gene; U1567 gene; U1568 gene; U1569 gene; U1570 gene; U1571 gene; U1572 gene; U1573 gene; U1574 gene; U1575 gene; U1576 gene; U1577 gene; U1578 gene; U1579 gene; U1580 gene; U1581 gene; U1582 gene; U1583 gene; U1584 gene; U1585 gene; U1586 gene; U1587 gene; U1588 gene; U1589 gene; U1590 gene; U1591 gene; U1592 gene; U1593 gene; U1594 gene; U1595 gene; U1596 gene; U1597 gene; U1598 gene; U1599 gene; U1600 gene; U1601 gene; U1602 gene; U1603 gene; U1604 gene; U1605 gene; U1606 gene; U1607 gene; U1608 gene; U1609 gene; U1610 gene; U1611 gene; U1612 gene; U1613 gene; U1614 gene; U1615 gene; U1616 gene; U1617 gene; U1618 gene; U1619 gene; U1620 gene; U1621 gene; U1622 gene; U1623 gene; U1624 gene; U1625 gene; U1626 gene; U1626.5 gene; U1627 gene; U1628 gene; U1629 gene; U1630 gene; U1631 gene; U1632 gene; U1633 gene; U1634 gene; U1635 gene; U1636 gene; U1637 gene; U1638 gene; U1639 gene; U1640 gene; U1641 gene; U

AUTHORS McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.
TITLE Comparative sequence analysis of the long repeat regions and
adjoining parts of the long unique regions in the genomes of herpes
simplex viruses types 1 and 2
JOURNAL J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)
MEDLINE 92113549
REFERENCE 4 (bases 1 to 154746)
AUTHORS Barnett,B.C., Dolan,A., Telford,E.A., Davison,A.J. and McGeoch,D.J.
TITLE A novel herpes simplex virus gene (UL49a) encodes a putative
membrane protein with counterparts in other herpesviruses
J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)
JOURNAL 5 (bases 1 to 154746)
MEDLINE 92356101
REFERENCE Dolan,A.
AUTHORS Direct Submission
TITLE Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church
Street, Glasgow, G11 5JR, UK
JOURNAL Location/Qualifiers
FEATURES
source 1..154746
/organism="human herpesvirus 2"
/strain="HG52"
/db_xref="taxon:10310"
misc_feature 1..254
/note="a sequence"
255
/note="start of TRL (terminal copy of Long Repeat region)"
440..934
/gene="RL1"
/number=1
440..126810
/gene="RL1"
join(440..934,1089..1379)
gene 1089..1379
/gene="RL1"
/codon_start=1
/protein_id="CAB06759.1"
/db_xref="PID:e308397"
/db_xref="PID:g1869821"
/db_xref="GI:1869821"
misc_feature 1089..1379
/db_xref="SWISS-PROT:P28283"
/translation="MSRRGPRRRGPRRRPRPGAPAVPRPGALPTADSDM
VPAYDSGAVESAPASSLRRMLVPOADSDDDADYAGNDAAEMANSPPSGGSKAP
EAPHAAPAAACPPPPRKEGPOPLRLPRLTTEYLARLSIRRRRPPASPAD
APRGVCSPPVQVHLYAVETAAALARGSGARARARDRRRRVAAAEAVTGPCLE
PEARARARARAHEDGGPAEEEAARSSAAAGGRRAY"
repeat_region 446..460
/rpt_family="7"
repeat_region 461..547
/rpt_family="6"
repeat_region 943..1070
/rpt_family="5"
exon 1089..1379
/gene="RL1"
/number=2
/product="neurovirulence factor"
1738..1743
/gene="RL1"
2303..124947
/gene="RL2"
2303..2377
/gene="RL2"
/number=1
join(2303..2377,2785..3462,3644..5368)
CDS
/gene="RL2"
/codon_start=1
/protein_id="CAB06760.1"
/db_xref="PID:e304156"
/db_xref="PID:g1869832"
/db_xref="GI:1869822"
/db_xref="SWISS-PROT:P28284"
/translation="MEPRGTSRADPPEPRPPROTPTQAPAPAHAMGLNDMDMLAS
SDSEETEVIGISDDLHSDSTSEAGSTDEMEAGLMDAATPPAPAPARSGSPAD
AOGSGGGPVEEAEAGGGGDDVCAVCTDELAPPLRQOSFCLHPFCIPCKMTPLR
NTCEPCNTPVAVLIVGVAAGSGSFITPVNDPRTRVEAAVRAAGTAVDFWTGNPRI

APRSLSGHTVRALSPTPMPGTDDEDDLDADYVYPAPRAPAPREGGAGATRG
SQPAPTRAPGAPARRSSSGGAPRACGSSGGGAPVAAYVPVATSPAPAGGRAG
ARRVEDAAAEGRTPPAROPRAOEPPIVSDSPSPRPAPGPISEFSSSAQY
SSGGGGGLPQSSSGRAPAVAPRVASPPRAAAPVSAADAAAGAPAVVDAA
RAPSMTAOQTDQAOISLGRATDARGSGPGAEGPGVPRGTNTFGAPHAAGA
AAPRRRGSDSGPAASSAASSAPRPLAPGVGAKRAAPRAPDSDGDRGGL
APAGAAPAPSPSSQAAYAAASSAASSAASSAASSAASSAASSAASSA
SSSAGAGGSVYASGAGERETSLGPRAPAPRPRCKARTRIAEAGPEGADPAP
GLTRYLPLAGVSVVALAPVYKTVTDCLPLDMEGHGAYVVLVDQGTGVADLLR
AAAPAWSRRLTLPKHARNCVPRPDYPPPAEWMNSLWMTPVGNMLFDGTLVADLFH
GLRSHHWSRQAGAPADGAPAGGE"
exon 2785..3462
/gene="RL2"
/number=2
3468..3607
/rpt_family="4"
exon 3644..5368
/gene="RL2"
/number=3
/function="modulator of cell state and gene expression"
4655..4771
/rpt_family="3"
complement(4811)
/gene="RL2"
/note="proposed LAT splice acceptor site"
5618..5623
/gene="RL2"
complement(7025)
/gene="RL2"
/note="proposed LAT splice donor site (5' end of stable
LAT species)"
complement(7732)
/gene="RL2"
/note="primary transcription initiation site for latency
associated transcripts (LAT's)"
9047..9193
/rpt_family="2"
9276..9297
/rpt_family="1"
9298
/gene="RL2"
/note="start of UL (Long unique region)"
9427..10101
/gene="UL1"
9427..10101
/gene="UL1"
/codon_start=1
/product="viral glycoprotein L"
/protein_id="CAB06761.1"
/db_xref="PID:e304262"
/db_xref="PID:g1869823"
/db_xref="GI:1869823"
/db_xref="SWISS-PROT:P28278"
/translation="MGFVCLFGLVYMGAMGSGQATEYVLSVIAKENVGIDITRVC
WRTPADVSMRTEAPSVIDYARIDGIFLRYPGDLDFLMDRAHRAVYINPFLFAG
FLIEDLSHVPADQETTRALRYKEIRDALGSRKQVSHAPVAGCVNPFYSSTRRC
VGRDLRPANTSTWEPVSSDDSAQSXPPLATQPVVALSNAPPRVSPTRRRRH
TRLRRN"
10211..10978
/gene="UL2"
10211..10978
/gene="UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="CAB06762.1"
/db_xref="PID:e304157"
/db_xref="PID:g1869824"
/db_xref="GI:1869824"
/db_xref="SWISS-PROT:P28275"
/translation="MFSASTPEPTGLSDADNPPLTSPVLDMAAFRAFLIDDAMR
PLEEPELANLTLTALAEDVROTEVLPREDVSWTYXCIPDDVRYVITIDQPH
HPOGAGHLASVRADVPFSLRNVLAANKTFDAMWSRGCUENARGVLLANT
LTVKGAASHSKLGMDFVGGVAVVLAARRPGLVFMWGAHONAIAPPRRHYVYIK

CDS
/number=1
/join(2303..2377,2785..3462,3644..5368)
/gene="RL2"
/codon_start=1
/protein_id="CAB06760.1"
/db_xref="PID:e304156"
/db_xref="PID:g1869822"
/db_xref="GI:1869822"
/db_xref="SWISS-PROT:P28278"
/translation="MEPRGTSRADGPPRRPPTGTPAPAHMGLINDQWLAS
SSEETEVGIISDDLDHROSTSEAGSTDEMFAGLMDAATPPARPAERKQSPAD
AQCSCGGPVEEABAGGGGDCVACTDEIAPLCQSPCLHPCIPCMKMLIR
NTPCLNTPAYALIVGTASGSESTPIYNDPRTVEAEAAVAGTAVDITNGNRT
APSLSLGHTVRLSPTPWPGTDEDDDLADVDVPPARAPRRARAGGGAATGCT
SQPATRPAPRRPSSSGGALIRAGVSGSGGCAVAAYVRVSLPPRAGGGAQ
ARVGGDAAAEGRTPPARQPAOEPIYISDPPSPRPGPGPLSFVSSSAQV
SSQGGGGGLQSSGGAAPRAAAYAPRSPRAAAYVSAADAAGPAPVADAH
RAVSRMTQAQDTQASLRAGATDAGSGGAGGAGPGVPGTNTPGAPAAAGA
AAPRRRGSDGPAASSASSAAPRSPPLAPGVAKRAAPRRAPRRDGGDHPL
APASAGAAPSPASOAAVAAASSASSASSASSASSASSASSASSASSASSA
SSAGAGGSYASASGERRRSLGPRAPAPRPGKCAKITHAGGPPRANRDPAP
GLRIPLIAGVSVVALAPYVNTVGTGDLPLDMETGIGATVVLVDGTGNADLLR
AAPAVSRRLLEHARNVCYRPDPYTPPASEWNSIMTVPVGNMLFDQGLVGLDFH
GLSRHPMSREOGAPAPAGDAPAGHGE"
2785..3462
/gene="RL2"
/number=2
3468..3607
/rpt_family="4"
3644..5368
/gene="RL2"
/number=3
/function="modulator of cell state and gene expression"
4655..4771
/rpt_family="3"
complement(4811)
/gene="RL2"
/note="proposed LAT splice acceptor site"
5618..5623
/gene="RL2"
complement(7025)
/gene="RL2"
/note="proposed LAT splice donor site (5' end of stable
LAT species)"
complement(7732)
/gene="RL2"
/note="primary transcription initiation site for latency
associated transcripts (LAT's)"
9047..9193
/rpt_family="2"
9216..9297
/rpt_family="1"
9298
/gene="RL2"
/note="start of UL (Long Unique region)"
9427..10101
/gene="UL1"
9427..10101
/gene="UL1"
/codon_start=1
/product="viral glycoprotein L"
/protein_id="CAB06761.1"
/db_xref="PID:e304262"
/db_xref="PID:g1869823"
/db_xref="GI:1869823"
/db_xref="SWISS-PROT:P28278"
/translation="KGFVCLFGLVYVANGAMGSGATERYVLSVTAKEVGLIRYC
MRPADDVSWRTAPSVIDARIDGILFRHCGDLPDPLMDRHAQRAYLVNPLFAG
FLEDDLSHVSVPADOTETTRALYKIRDALGSRKQAVSHAPVACGVNEDSRTRR
VGRDLRPANTSTWEPVPSDDASSOSKPLATOPVLAISNAPRRVSPTRGRRH
TRLRRN"
10211..10978
/gene="UL2"

CDS
10211..10978
/gene="UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="CAB06762.1"
/db_xref="PID:e304157"
/db_xref="PID:g1869824"
/db_xref="GI:1869824"
/db_xref="SWISS-PROT:P28275"
/translation="MFSATPEQPLGSGDATPLPVSPLDMAAFRRALIDAMR
PLIEPLANPLTARLIAEYDRRCQTEVYPPREDVFSWTRYPDVRVILIGOPYH
HGOAGLAFSVRAVPVPSLRNLYAANKCYPARMSRGCLERKADGVLLNTT
LVKRGAAASHPGKIGDRVGVVORLARREGIVFMGMGAINAQNARIPRQHYLK
FHSRPLSVPRGTGCOHFLAANRYLETDRIMEIDMSV"
11019..11024
/gene="RL2"
11033..11734
/gene="UL3"
11033..11734
CDS
Query Match 16.1%; Score 63; DB 17; Length 154746;
Best Local Similarity 48.7%; Pred. No. 0.083;
Matches 171; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 37 GCCCTGCGCTGGCCAGAGCGCGGAGCCGAGACCCCGCTCGGTGAGACTGGGGGTGCA 96
Db 148472 GCCGGGGGAGCCGGCCGGGGGAGCGGCCGGGGGAGCGGCCGGGGGAGCGGGGG 148413
QY 97 GGTGCCGAGACCTTACCCAGCGCGGGAGTACCTCCCGCTCACACCTCGGCTGCAGTTTC 156
Db 148412 GACGGGCGGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148353
QY 157 CCTGGTGCCCGCCGAGACGCTGCCCCGCGCTGAGAGGATGCGGGGAGCGGGGAGCGGGG 216
Db 148352 GCCGGGGGCGCGGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148293
QY 217 GGGGGGGGGGCTCTGACGTGAGAGAGCGCGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 276
Db 148292 GGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148233
QY 277 CGGCTCCTTAAAGCGCGGAGCGCGGAGCGGAGCGGTGCTCTGTGAGCCGAGCGGAGCGG 336
Db 148232 GACGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148173
QY 337 CCGGGGAGACCGGAGACGTGAGCGCGCGCTCGCTGACGTCCCGGAGC 387
Db 148172 CGGGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGGAGCGGGG 148122
RESULT 12
CH19R30879/c
LOCUS CH19R30879 41936 bp DNA PRI 01-APR-1997
DEFINITION Homo sapiens DNA from chromosome 19-cosmid R30879 containing USR2,
genomic sequence.
ACCESSION AD000684
NID AD000684
VERSION 91905917
KEYWORDS AD000684.1 GI:1905917
SOURCE chromosome 19; transcription factor.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 41936)
AUTHORS Lamerdin,J.E.
TITLE Direct Submision
JOURNAL Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA, USA, 94551 janefacg.1lnl.gov owetornak.1lnl.gov
COMMENT
GSDS:S:1010600.
Human Genome Center
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
7000 East Avenue


```

exon      5303..5390
           /gene="USF2"
           /number=6
Intron    5391..5667
           /gene="USF2"
           /number=6
exon      5668..5726
           /gene="USF2"
           /number=7
Intron    5727..13283
           /gene="USF2"
           /number=7
exon      13284..13378
           /gene="USF2"
           /number=8
Intron    13379..13531
           /gene="USF2"
           /number=8
exon      13532..13660
           /gene="USF2"
           /number=9
Intron    13661..13752
           /gene="USF2"
           /number=9
exon      13753..14380
           /gene="USF2"
           /number=10
BASE COUNT 2766 a 4210 c 4353 g 3110 t 1 others
ORIGIN

```

```

Query Match      15.7%   Score 61.6; DB 10; Length 14440;
Best Local Similarity 53.1%   Pred. No. 0.29;
Matches 155; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

```

```

QY  55 GCGCGGAGGCGCCAGAGCCCGCTCGGTGAGAGACTGGGGGTGAGAGTCCCGGAGCGTACCC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3807 GCGCGGAGGCGCCAGAGCCCGCTCGGTGAGAGACTGGGGGTGAGAGTCCCGGAGCGTACCC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  115 AGCGCGGAGAGTACTCCCGCTACACCTCGGCTGACATTCTCGGTGAGAGTCCCGGAGCGTACCC 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3747 GCGCGGAGGAGAGCGGTGCGCGAGGAGCGGAGTCCAGACCC---GGGTCCAGAGATGTC 3691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  175 CGCTGCGCCCGGCTGAGAGGATGCGGGGCGGAGCGGGGCGGGGCGGCTGCTGTCAC 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3690 CATGGGAGGAGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGAGGAGGAGGAG 3631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    235 GTGAGAGAGGCGCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAAGCGCG 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3630 GAGGAGGAGGAGGAGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCTCTTAAAGCGCG 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  295 CGAGCCGAGCGGCGGAGGTCTCTGTGCGCGAGGCGGAGCGCGCGCGGAGCA 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3570 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 3519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
LOCUS      SCE94      38532 bp      DNA
DEFINITION Streptomyces coelicolor cosmid E94.
ACCESSION AL049628
VERSION    94585581
KEYWORDS   AL049628.1 GI:4585581
SOURCE     Streptomyces coelicolor.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Streptomyces coelicolor.
            Actinomycetales; Streptomycetaceae; Streptomyces.

```

REFERENCE
AUTHORS Redenbach, M., Kleser, H. M., Denapate, D., Eichner, A., Cullum, J., Kinash, H. and Hopwood, D. A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 38532)
AUTHORS Oliver, K. and Harris, D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 38532)
AUTHORS Bentley, S. D., Parkhill, J., Barrell, B. G. and Rajandream, M. A.
TITLE Direct Submission
JOURNAL Submitted (12-APR-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)
CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., (Gene 30:157-66(1984)) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid E94 lies between E7 and E126 on the AseI-E genomic restriction fragment.

FEATURES
SOURCE location/Qualifiers
1..38532
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid E94"
complement(1..1512)
/gene="SCE94.01c"
1..105
/gene="SCE94.01c"
/note="Nominal overlap with Streptomyces coelicolor E7."
complement(<1..1512)
/gene="SCE94.01c"
/note="SCE94.01c, partial CDS unknown, alanine-rich protein, len>503aa"
/codon_start=11
/transl_table=11
/label="SCE94.01c"
/product="alanine-rich hypothetical protein"
/protein_id="CAB40850.1"
/db_xref="PID:e1424928"
/db_xref="PID:g4585582"


```

/db_xref="GI:4585582"
/translation="MRSSPRTTPTTGAAHRAHREARDRTAARALAPPARPAA
VDAYLDGLFTYCLSLCDHDAATATAGVLAIAERGRGRAPADRSMLIALARNA
CIRKAEKOKRPSTHAGRPAPVTRQATTPATPTTINDETHORHRSERLALAMP
EAAGTTPQREALELAVRHHLAPHEVAALVGLDIAATRELLASACEVERERFALAV
ETGCAGAVARLGLNARYLSTALRELVHVDCCRCRTAERAIIPGRWPGSTVPAE
LPVEAPRTALVLAHATSRASAPRDFRPFMDPRAARDRLRAPATTTVA
TVAAVAVLMAAYRGTPVVEGEGRASAPADPDADGSGAAGVEMAGNAST
TPGNGFGKGGADVSEVYGVSGAGSKAAGLETVAGDGTDTTTLTATGAPYHWS
ASVGASMLTSSGSLTRPGEAVTVRVTDHLREPSGMSARVAVSPAGAVAT"
2126..3535
/gene="SCE94.02"
2126..3535
/gene="SCE94.02"
/note="SCE94.02, possible DNA repair protein, len: 469aa;
similar to many eg. SW:RADA_ECOLI Rada, DNA repair protein
(mechanism unknown) from Escherichia coli (460 aa) fasta
scores: opt: 1218, z-score: 1309.3, E(): 0, (41.4%
identity in 452 aa overlap). Contains Prosite match to
PS00017 ATP/GTP-binding site motif A (P-loop)."
```

```

/note="SCE94.04c, unknown, len: 126aa"
/codon_start=1
/translation="MRSSPRTTPTTGAAHRAHREARDRTAARALAPPARPAA
VDAYLDGLFTYCLSLCDHDAATATAGVLAIAERGRGRAPADRSMLIALARNA
CIRKAEKOKRPSTHAGRPAPVTRQATTPATPTTINDETHORHRSERLALAMP
EAAGTTPQREALELAVRHHLAPHEVAALVGLDIAATRELLASACEVERERFALAV
ETGCAGAVARLGLNARYLSTALRELVHVDCCRCRTAERAIIPGRWPGSTVPAE
LPVEAPRTALVLAHATSRASAPRDFRPFMDPRAARDRLRAPATTTVA
TVAAVAVLMAAYRGTPVVEGEGRASAPADPDADGSGAAGVEMAGNAST
TPGNGFGKGGADVSEVYGVSGAGSKAAGLETVAGDGTDTTTLTATGAPYHWS
ASVGASMLTSSGSLTRPGEAVTVRVTDHLREPSGMSARVAVSPAGAVAT"
2126..3535
/gene="SCE94.02"
2126..3535
/gene="SCE94.02"
/note="SCE94.02, possible DNA repair protein, len: 469aa;
similar to many eg. SW:RADA_ECOLI Rada, DNA repair protein
(mechanism unknown) from Escherichia coli (460 aa) fasta
scores: opt: 1218, z-score: 1309.3, E(): 0, (41.4%
identity in 452 aa overlap). Contains Prosite match to
PS00017 ATP/GTP-binding site motif A (P-loop)."
```


subunit.
ACCESSION 212168
ID 9311336
VERSION 212168.1 GI:311336
KEYWORDS GTP binding protein; GTP binding protein alpha subunit.
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1927)
AUTHORS Ishikawa, Y. and Homcy, C.J.
TITLE cDNA sequence for alpha subunit of stimulatory guanine nucleotide
binding protein from canine heart
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1927)
AUTHORS Ishikawa, Y.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1992) Yoshihiro Ishikawa, Department of
Pharmacology, Columbia University, 630 W, 168th St., New York, NY,
10032, USA

FEATURES
source Location/Qualifiers
1..1927
/organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="Ventricular heart muscle"
/clone="Clone 7M"
354..1538
/codon_start=1
/product="stimulatory GTP binding protein"
/protein_id="CAA78161.1"
/db_xref="pid:311337"
/db_xref="gi:311337"
/db_xref="swiss-prot:P04895"
translation="MGCLGNSKTEDQNRNEKAQREANKRIEKOLOKQVYRATHRL
LLGAGESGKSTIVKQRIILHVNGEGEEDPQARNSDGEKATVODIKNNLKEA
IETIVAAASNLPVELANPERQFVDYILISVNPDPDPPEFEHAKALMEDEGVR
ACERSNEQLIDCAQYFLDKDYKQADYPSDDLRCRVLTSGIFETKQYDKVN
FMFPGGGRDRERKRIQCFNDVTAITFVASSSTINVIREDNQTNRLOEALNFKST
WNNRLRTISVILFNKODLAELKYLAKSKIEDIEFPFARTTPEDATPEGDPYV
TRAKYFIRDEFIRISTASGDGRHYCPHFTCAVDTEINIRVFNDCRDIIOHMLROYE
LL"

BASE COUNT 493 a 576 c 500 g 358 t
ORIGIN

Query Match 15.6%; Score 61; DB 3; Length 1927;
Best Local Similarity 52.1%; Pred. No. 0.69;
Matches 159; Conservative 0; Mismatches 145; Indels 1; Gaps 1;

QY 31 TGGGCAAGCCCTGCGCTGCGCCAGAGCGCGAGAGCCCGCTCGGTGAGACTGGG 90
DB 319 TGGGCGGCGAG 260
QY 91 GGTGAGAGTGTCCCGAGAGCGTACCCAGC-GCCGGAGTACCTCCCGCTCACACTGGGCT 149
DB 259 GGGCGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCTCGGGGA 200
QY 150 GCAGTTCCTGGGTGGCGCGCGAGACCTGCGCGCGCGCTGAGGAGATGGCGGGCGGGGA 209
DB 199 GAGAGCGGTGCGTA 140
QY 210 CGGGGCGCGGGGCGCGGCTGCTACGTGAGAGAGCGCGCGGGCGCGGGCGGGGCGG 269
DB 139 CCGAGGGGAGCGCGGAGCTCCCGCCCTCGAGCCGAGGCGGAGGGGCGCTGATGCCCGC 80
QY 270 GCGCGCCGCGCTCTTAAGAGCGCGGAGCGCGAGCGCGAGAGTGCCTGTGGCGCGCAG 329
DB 79 CGCGGGCGCGAGCGGACCAAGAGCAGAGACTGCGGGGAGACTGCTGCCCGCGCGCGC 20
QY 330 CGCAG 334
DB 19 CGCGG 15

Search completed: September 17, 1999, 22:05:04
Job time: 16487 sec

1	72	18.4	1876	1	T48339
2	72	18.4	1876	1	V12197
3	69.2	17.7	3851	1	Q51543
4	64	16.3	3198	1	X02974
5	63.6	16.2	114955	1	X53491
6	60.6	15.5	1281	1	Q23396
7	57.8	14.7	11830	1	V18130
8	56.4	14.7	114955	1	X53491
9	56.4	14.4	12001	1	Q76213
10	56	14.3	1130	1	V60339
11	52.4	13.4	12001	1	Q76213
12	51.8	13.2	4403	1	Q47927
13	51.8	13.2	2218	1	Q47929
14	51.8	13.2	4257	1	V10362
15	51.8	13.2	4257	1	V68520
16	51.2	13.1	5228	1	V81884
17	50.4	12.9	7193	1	V50431
18	50.2	12.8	1335	1	Q23295
19	49.6	12.7	1303	1	Q23297
20	49.6	12.7	407	1	V44330
21	49.6	12.7	407	1	V64539
22	49.2	12.6	701	1	Q76209
23	49	12.5	4405	1	Q25444
24	49	12.5	4405	1	T05556
25	49	12.5	223	1	T66412
26	49	12.5	223	1	V01700
27	48.4	12.3	7011	1	V20464
28	48.4	12.3	801	1	V73801
29	48	12.2	4070	1	T91861
30	48	12.2	117213	1	V62176
31	47.8	12.2	1032	1	Q44281
32	47.6	12.1	6225	1	X55273
33	47.4	12.1	1610	1	Q57657
34	47.4	12.1	1608	1	T72547
35	47.4	12.1	4425	1	V32841
36	47.2	12.0	117213	1	V62176
37	46.8	11.9	4448	1	T35520
38	46.8	11.9	339	1	V44425
39	46.8	11.9	339	1	V64534
40	46.2	11.8	4356	1	Q37543
41	46.2	11.8	4356	1	Q95540
42	46	11.7	1725	1	Q11243
43	46	11.7	1894	1	V08416

PA (SIBI-) SIBIA NEUROSCIENCES INC.
(SIBI-) SIBIA NEUROSCIENCES INC. ASSOC
(SIBI-) SIBIA NEUROSCIENCES INC. ASSOC
(SIBI-) SIBIA NEUROSCIENCES INC. ASSOC

CC myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 3198 BP; 807 A; 847 C; 681 G; 846 T;

Query Match 16.3%; Score 64; DB 1; Length 3198;
Best Local Similarity 51.3%; Pred. No. 0.0022;
Matches 176; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 26 GGGCTGGGGAGAGCCCTGGCCGACAGAGCGCCAGACCCCGCTGGTGAGAG 85
Db 3132 GGG 3073
QY 86 CTGGAGGTGAGAGTCCCGGAGACCTACCCAGCGCCGAGTACCTCCGCTACACCTCG 145
3072 GGG 3013
QY 146 GAGTCACTTCCTGGGTGGCCGCGAGACGCTGAGCGGCTGGAGGATGCGCGGCG 205
3012 GGG 2953
QY 206 GGGAGCGGG 265
Db 2952 GGG 2899
QY 266 GGGGCGGG 325
Db 2888 GGG 2839
QY 326 CAGGCGCAGGCGCCGCGCAGACCGAGAGCTGAGAGCGCGCGCG 368
Db 2838 GCGGCGCCCGCCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2796

RESULT 5

ID X53491 standard; DNA; 114955 BP.

AC X53491;
DT 05-JUL-1999 (first entry)
DE Human adenovirus A1 receptor antisense oligonucleotide fragment.
KW Antisense oligonucleotide; multiple target; antisense treatment;
KW impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
KW respiratory distress syndrome; pain; cystic fibrosis;
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
OS Synthetic.
PN WO913886-A1.
PD 25-MAR-1999.
PF 17-SEP-1998; U19419.
PR 09-JUN-1998; US-093972.
PR 17-SEP-1997; US-059160.
PA (UPEC-) UNIT EAST CAROLINA.
PI NYCE JW;
DR WPI; 99-229400/19.
PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction
PT Disclosure; Page 37, 120pp; English.
CC The specification describes antisense oligonucleotides (X52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene
CC initiation codons, genomic flanking regions, intron-exon borders, the
CC 5'-end, the 3'-end and the junction between coding and non-coding
CC regions and all segments of RNAs encoding proteins associated with one
CC or more diseases, conditions or mixtures. The antisense oligonucleotides
CC may be derived from sequences X55272-74. These multiple target

CC oligonucleotides (specifically X55180-271) can be used for the antisense
CC treatment of diseases and conditions. Typical diseases and conditions
CC are those associated with impaired respiration and inflammation,
CC including lung diseases, pulmonary vasoconstriction, inflammation,
CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC metastases, as well as all types of cancers which may metastasize or have
CC metastasized to the lungs, including breast and prostate cancer.
SQ Sequence 114955 BP; 6071 A; 2947 C; 3672 G; 2158 T;

Query Match 16.2%; Score 63.6; DB 1; Length 114955;
Best Local Similarity 38.5%; Pred. No. 0.0013;
Matches 130; Conservative 27; Mismatches 179; Indels 2; Gaps 1;

QY 33 GGCAGCCCTGGCTGGCCAGAGGCGGAGGCGGAGACCCCTCGGTGAGACTGGGGG 92
Db 108466 GGG 108525
QY 93 TGAAGTCCCGGAGAGTACC--CAGCGCGGAGATACCTCCGCTACACCTGGGCTG 150
Db 108526 CGGG 108585
QY 151 CAGTTCCTGGGTGG 210
Db 108586 CGGG 108645
QY 211 GGG 270
Db 108646 SNNNDNNGCGCGGG 108705
QY 271 GCGCGCGGGCTCTTAAAGGCGCGGAGCGGAGCGGAGGTGCTGTGTGCGGAGGC 330
Db 108706 GCGCGCGGCGSNNNDNNGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 108765
QY 331 GCAGGCGCGGGCGACACCGGAGAGTGGAGGCGGCGCGG 368
Db 108766 GGG 108803

RESULT 6

ID Q23296 standard; DNA; 1281 BP.

AC Q23296;
DT 28-JUL-1992 (first entry)
DE HSV-1 (MGH-10) ICP34.5 gene.
KW Herpes simplex virus; avirulent; vaccine; immunisation; ss.
OS Herpes simplex virus.
FH Key
FT cds
FT Location/Qualifiers
FT 134..893
FT /*tag- a

PN WO9204050-A.
PD 19-MAR-1992.
PF 10-SEP-1991; U06532.
PR 10-SEP-1990; US-579834.
PA (ROIZ/) ROIZMAN B.
PI Roizman B;
DR WPI; 92-114074/14.
DR P-PSDS; R21707.
PT New recombinant Herpes Simplex Virus vaccines - rendered
PT avirulent by deletion of ICP 34.5 gene encoding active gene prod.
PT Disclosure; Fig 1; 78pp; English.
CC Herpes simplex virus-1 (MGH-10) was isolated from a recurrent facial
CC vesicle and passaged a maximum of four times in cells in culture.
CC Viral DNA was prep'd. from virions that accumulated in the cytoplasm
CC of infected VERO cells (Kieft et al., J. Virol., 8, 125-132 (1971)).
CC The BamHI SP junction fragments conty. the domain of the gene that
CC specified ICP34.5 were cloned into a pUC18 plasmid using HSV-1(F)
CC sequences as probes in colony blot hybridisation. Further

	Query Match:	14.7%	Score 57.8;	DB 1,	Length 11820;
	Best Local Similarity	54.6%;	Pred. No. 0.017;		
	Matches 183; Conservative	0;	Mismatches 142;	Indels 10;	Gaps
QY	14 GAGGTAGAGCCTTGCCCTTGGGCGACGCCCTCTGAC--CTGACCAGAGAGCGCGAGCCGAGAGC	71			
Db	1246 GCGGCCTGATGCGATCATTTGGCGGCCCCCGGCTMGACAGCGCGGGGGCGGGCGGGCGGCG	1187			
QY	72 CCGETTCGATTGAGACTGAGGGGTGGAGGTGCCCGGAGCGTACCGACGCCGGGAGTACTTC	131			
Db	1186 CGGGCGCGGGGGGACGCGCGCGGGCGGG-----GGATTTCACCTGGCGCGCGCGACGACGG	1132			
QY	132 CCGGTACACACTCGGGGTGGAGTTCCTCGGTGGCGCGCGCACAGACGCTGGCGCGGCTGGA	191			
Db	1131 TGCGCCGAGGACAGCGGGGATCCAGACC---GGTTCACACATGCTCATGGGGGGGGCGGGG	1075			
QY	192 GGATGTGCGGGGCGGGACGAGGGCGGGGCGGGGCTCTACGTGAGAGAGCGCGCGG	251			
Db	1074 CGGGGGGGGGCGCGGGCGCGCGGGGGGAGGGAGGAGGAGGGAGGAGGAGGAGGAGGAGG	1015			
QY	252 GCGGCGCGGGCGGGGCGCGCGCGCGCTCTTAAAGCGCGCGAGCCGAGCGGAG	311			
Db	1014 CGGCGCGCGCGGGGGGCGCGCGAGCCCGGCTCAACGCGCGCGGAGAGAGGAGCGAGG	955			
QY	312 TGCCTCTGTGGCCCGCAGCGCCAGGCCCCGGCGACA	346			
Db	954 AGATACGGAGCCGCTCGCGCTGATCACGGGGACA	920			
RESULT	8				
ID	X53491/c				
AC	X53491 standard; DNA; 114955 BP.				
DT	05-JUL-1999 (first entry)				
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.				
KW	Antisense oligonucleotide; multiple target; antisense treatment;				
KW	impaired respiration; inflammation; lung disease;				
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;				
KW	acute asthma; allergy; asthma; impeded respiration;				
KW	respiratory distress syndrome; palm; cystic fibrosis;				
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;				
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;				
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;				
KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;				
CS	prostate cancer; ss.				
SC	Synthetic.				
NC	MO9913886-A1.				
PF	25-MAR-1999.				
PR	17-SEP-1998; U19419.				
PR	09-JUN-1998; US-093972.				
PR	17-SEP-1997; US-059160.				
PA	(UYEC-) UNIV EAST CAROLINA.				
PI	Nyce JM;				
PT	WPI; 99-229400/19.				
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary				
PT	vasoconstriction				
PS	Disclosure; Page 37; 120pp; English.				
CC	The specification describes antisense oligonucleotides (X52869-X55271)				
CC	directed against at least 2 mRNAs selected from target genes, coding and				
CC	non-coding regions of RNAs corresponding to target genes, gene				
CC	initiation codons, genomic flanking regions, intron-exon borders, the				
CC	regions, the 3' end and the junction between coding and non-coding				
CC	regions and all segments of RNAs encoding proteins associated with one				
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides				
CC	may be derived from sequences X55272-74. These multiple target				
CC	oligonucleotides (specifically X55180-271) can be used for the antisense				
CC	treatment of diseases and conditions. Typical diseases and conditions				
CC	are those associated with impaired respiration and inflammation,				
CC	including lung diseases, pulmonary vasoconstriction, inflammation,				
CC	allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,				
CC	respiratory distress syndrome, pain, cystic fibrosis, pulmonary				
CC	hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive				
CC	pulmonary disease (COPD), and cancers such as leukemias, lymphomas,				

```
CC carcinoma e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC metastases, as well as all types of cancers which may metastasize or have
CC metastasized to the lungs, including breast and prostate cancer."
SQ Sequence 114955 bp; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match      14.7%; Score 57.8; DB 1: Length 114955;
Best Local Similarity 35.0%; Pred. No. 0.011;
Matches 131; Conservative 45; Mismatches 192; Indels 6; Gaps 1.

OY    22 GCCTGACCTTGGGCAGCCCCCTGGCGCAGAGGCCGAGCCGACTCGTCG 81
      ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db 104995 GCCCGGCGCGCGCGCGCGCGCGCGCCVGNHHNNNSGCGCCGCGCGCGCGCGCGCGC
OY    82 GAGACTGGGGGTGAGATGTCCCGGAGCGGTACCAGCGCGCGGAGTA-----CTCCCGC 135
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104935 CVGCGNHHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCN 104876
OY    136 TCACACCTTGCGGCTCACTTCCCTGGGTGGCCCGCGAGACGCTGCGCGCGGCTGAAGGA 195
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104875 HHNNNSVGGCCVCGCGNHHNNNSCVGGCCGCGNHHNNNSCCVGGCCVCGCGNHHNNSC
OY    156 TGCGCGGCGCGGAGCGGGGCGCGGGCGGGGCGGTCTCACGTGAAGAGCGCGCGCGCGCG 255
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104815 CCVGGCCVCGCGGNNHHNNNSGCCVGGCCVCGCGCGNHHNNNSCCGCCVGGCCVCGCGNHNH
OY    256 GGCGGGGCGGGGCGCGCGCGCGCGCGCTCTTAAGGCGCGCGAGCGCGGAGGAGGTGCC 315
      : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 104755 NSGCGCCCVGGCGCGCGCGGNNHHNNNSCGCGCCCGCGCGCGCGNHHNNNSGCCCGCCVGG
OY    316 TCTGTGGCCGCGAGCGCAGCGCGCGCGCGAGCACGAGACTGAGAGCGCGCGCGCTCGTG 375
      : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 104635 CCVGGGCGNHHNNNSGCGCGCGCGCGCGCGCGCGCGCGNHHNNNSCGCGCGCGCGCGCGG
OY    376 CAGCTCCGCGGACTC 389
      : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 104635 NNHHNNNSCGCGCGC 104622

RESULT_9
ID O76213/C
AC O76213 standard; DNA: 12001 BP.
DR O4-AUG-1995 (first entry)
DE HSV L/ST region.
KW HSV, junction-spanning transcript; L/ST; therapy; virucide; ds.
OS Herpes simplex virus type 1.
FH Key Location/Qualifiers
FT tata_signal 199..204 /tag= a
FT misc_binding 220..225 /tag= b
FT misc_rna 229 /note= "ICP4 binding site"
FT /tag= c
FT /note= "5' end of the L/STs"
FT misc_rna 370..372 /tag= d
FT /tag= d
FT /note= "first codon of ORF-1"

PN WO9428156-A.
PD 08-DEC-1994.
PE 20-MAY-1994: U05770.
PR 20-MAY-1993: US-065146.
PA (DAND ) DANA FARBER CANCER INST INC.
PI Schaffer PA, Yen L;
PT WPI: 95-022825/03.
PT Herpes Simplex Virus (HSV) specific junction spanning transcript
PT -for inhibiting HSV L/ST synthesis, in the treatment of HSV
PT infection.
PS Disclosure: Page 38-44; 64pp; English.
AN An HSV-specific junction-spanning transcript (L/ST) maps at the 5'
end to the b repeat sequences of HSV DNA at approx. 3 and 125 kb,
CC
```


CC and extends into the c repeat sequences of HSV DNA. Compounds
CC that inhibit HSV L/ST synthesis may be used as anti-HSV
CC virucides. The HSV-1 DNA sequence in the region of the L/STs
CC is given in 076213
SQ Sequence 12001 BP; 1568 A; 4557 C; 4315 G; 1561 T;

Query Match 14.4%; Score 56.4; DB 1; Length 12001;
Best Local Similarity 50.3%; Pred. No. 0.028;
Matches 164; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

QY 43 GGCCTGCGCAGAGCGCGAGCGCGAGCGCGCTCGGTGAGACTGGGGGTGAGAGTCC 102
DB 2319 GGGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2260
QY 103 CGGAGCGTACCGAGCGCGCGGAGTACCTCCGCTCACACCTTCGGGCTCACTTCCCTG 162
DB 2259 CGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2200
163 TGGCCGCGGAGACGCTGCGCGGCTGAGGGATGCGGGGCGGGGAGCGGGGGGCG 222
2199 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2140
QY 223 GGGGCTGCTACGTGAGAGCGCGCGGGGGCGGGCGGGGGCGCGCGCGGCTC 282
DB 2139 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2080
QY 283 CTAAAGCGCGCGAGCGCGAGCGCGTCTGTGCGCGAGCGCGAGCGCGCGG 342
DB 2079 GAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGCGAGGGGGG 2021
QY 343 GACAGCGGAGCGTGTGAGCGCGCGCG 368
DB 2020 GAGGGGGCGAGGGGGCGAGGGGGCG 1995

RESULT 10
V60339
ID V60339 standard; cDNA; 1120 BP.
AC V60339; (first entry)
DT 04-DEC-1998
KW cDNA sequence of fibroblast growth factor-2 (FGF-2).
KW Fibroblast growth factor-2; FGF-2; leaderless protein; inhibition;
KW export; angiogenesis; restenosis; treatment; tumour; inflammation;
KW cell proliferation; diabetes; retinopathy; infection;
KW polycystic kidney disease; atherosclerosis; ss.
OS Homo sapiens.
MO9837880-A1.
03-SEP-1998.
25-FEB-1998; U03689.
26-FEB-1997; US-807014.
PA (CIBL-) CIBLEX CORP.
PI Baird A, Florkiewicz R;
FT WPI; 98-495377/42.
PT Inhibiting export of leaderless protein with agent that inhibits
PT binding to transporter protein - especially for treating
PT angiogenesis and restenosis by preventing export of fibroblast
PT growth factor; also methods for identifying leaderless proteins and
PT their transporters
PS Claim 2; Pages 53-54; 116pp; English.
CC The present sequence encodes fibroblast growth factor-2 (FGF-2), a
CC leaderless protein. A leaderless protein refers to a protein that is
CC found in an extracellular environment, but lacks a canonical leader
CC sequence. The specification describes a method for inhibiting export of
CC a leaderless protein from a cell. The method comprises treating
CC the cell with an agent that inhibits binding between the leaderless
CC protein and a transport molecule. Treatment with the inhibiting agent
CC is specifically used to treat angiogenesis and restenosis, i.e. where
CC expression of FGF-2 is inhibited, and the agent is applied to endothelial
CC or smooth muscle cells. Other applications are treatment of tumours
CC (melanoma, teratocarcinoma, ovarian carcinoma, bladder cancer and
CC neuroblastoma), inflammation, cell proliferation, complications of
CC diabetes (e.g. retinopathy), viral, bacterial or fungal infections,

CC polycystic kidney disease and atherosclerosis.
SQ Sequence 1120 BP; 260 A; 293 C; 342 G; 225 T;

Query Match 14.3%; Score 56; DB 1; Length 1120;
Best Local Similarity 47.7%; Pred. No. 0.052;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 41 CTGGCCCTGCGCAGAGCGCGAGCGCGAGCGCGCTCGGTGAGACTGGGGGTGAGAGTG 100
DB 83 GGGCTCTCCCAAGCGCGCGCTCCCGGAGACACCATCTGTAACCCCAAGTCCCGGGCG 142
QY 101 CCGGAGCGTACCGAGCGCGCGGAGTACCTCCGCTCACACCTTCGGGCTCACTTCCCTG 160
DB 143 CCGGCTGCGCGCGCGACCGAGGGGGCGCGGAGACAGAGAGGGGGCGAGCGGCTGAGGCTG 202
QY 161 GGTGGCGCGCGAGACGCTGCGCGGCTGGAGGAGTGGCGGGGGCGGAGCGGGGGG 220
DB 203 GGGGACCGCGCGCGCGCGCGCTGCGCGGCGGAGGAGGCTGGGGGGCGGGGGG 262
QY 221 GCGGGGCTGCTACGTGAGAGCGCGCGGGGGCGGGGGGGCGGGGGCGCGCGCGG 280
DB 263 CCGTGGCCCGAGCGCGGCTCGAGAGCGCGGGCGCGGGCGAGCGCGGCTCCCGCGG 322
QY 281 TCCTTAAGCGCGCGAGCGCGAGCGCGAGTGTCTGTGCGCGAGCGCGAGCGCGG 340
DB 323 CGGCTCCAGCGCGCTCGGGGATCCCGCGGGGGCGCGGAGGACCATGCGAGCGGGAGCA 382
QY 341 GCGAGACCGAGAGCGTGTGAGCGCGCGCGCTGCTGCACTCCG 384
DB 383 TCACACGCTGCTGCGCGCTGCGCGAGATGCGGCGAGCGCGG 426

RESULT 11
Q76213
ID Q76213 standard; DNA; 12001 BP.

AC Q76213;
DT 04-AUG-1995 (first entry)
DE HSV L/ST region.
KW HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.
OS Herpes simplex virus type 1.
FH Key location/Qualifiers
FT 199..204
/*tag- a
FT misc_binding 220..225
/*tag- b
FT misc_rna 229
/*tag- c
/*tag- d
/*tag- e
/*tag- f
/*tag- g
/*tag- h
/*tag- i
/*tag- j
/*tag- k
/*tag- l
/*tag- m
/*tag- n
/*tag- o
/*tag- p
/*tag- q
/*tag- r
/*tag- s
/*tag- t
/*tag- u
/*tag- v
/*tag- w
/*tag- x
/*tag- y
/*tag- z
/*tag- AA
/*tag- AB
/*tag- AC
/*tag- AD
/*tag- AE
/*tag- AF
/*tag- AG
/*tag- AH
/*tag- AI
/*tag- AJ
/*tag- AK
/*tag- AL
/*tag- AM
/*tag- AN
/*tag- AO
/*tag- AP
/*tag- AQ
/*tag- AR
/*tag- AS
/*tag- AT
/*tag- AU
/*tag- AV
/*tag- AW
/*tag- AX
/*tag- AY
/*tag- AZ
/*tag- BA
/*tag- BB
/*tag- BC
/*tag- BD
/*tag- BE
/*tag- BF
/*tag- BG
/*tag- BH
/*tag- BI
/*tag- BJ
/*tag- BK
/*tag- BL
/*tag- BM
/*tag- BN
/*tag- BO
/*tag- BP
/*tag- BQ
/*tag- BR
/*tag- BS
/*tag- BT
/*tag- BU
/*tag- BV
/*tag- BW
/*tag- BX
/*tag- BY
/*tag- BZ
/*tag- CA
/*tag- CB
/*tag- CC
/*tag- CD
/*tag- CE
/*tag- CF
/*tag- CG
/*tag- CH
/*tag- CI
/*tag- CJ
/*tag- CK
/*tag- CL
/*tag- CM
/*tag- CN
/*tag- CO
/*tag- CP
/*tag- CQ
/*tag- CR
/*tag- CS
/*tag- CT
/*tag- CU
/*tag- CV
/*tag- CW
/*tag- CX
/*tag- CY
/*tag- CZ
/*tag- DA
/*tag- DB
/*tag- DC
/*tag- DD
/*tag- DE
/*tag- DF
/*tag- DG
/*tag- DH
/*tag- DI
/*tag- DJ
/*tag- DK
/*tag- DL
/*tag- DM
/*tag- DN
/*tag- DO
/*tag- DP
/*tag- DQ
/*tag- DR
/*tag- DS
/*tag- DT
/*tag- DU
/*tag- DV
/*tag- DW
/*tag- DX
/*tag- DY
/*tag- DZ
/*tag- EA
/*tag- EB
/*tag- EC
/*tag- ED
/*tag- EE
/*tag- EF
/*tag- EG
/*tag- EH
/*tag- EI
/*tag- EJ
/*tag- EK
/*tag- EL
/*tag- EM
/*tag- EN
/*tag- EO
/*tag- EP
/*tag- EQ
/*tag- ER
/*tag- ES
/*tag- ET
/*tag- EU
/*tag- EV
/*tag- EW
/*tag- EX
/*tag- EY
/*tag- EZ
/*tag- FA
/*tag- FB
/*tag- FC
/*tag- FD
/*tag- FE
/*tag- FF
/*tag- FG
/*tag- FH
/*tag- FI
/*tag- FJ
/*tag- FK
/*tag- FL
/*tag- FM
/*tag- FN
/*tag- FO
/*tag- FP
/*tag- FQ
/*tag- FR
/*tag- FS
/*tag- FT
/*tag- FU
/*tag- FV
/*tag- FW
/*tag- FX
/*tag- FY
/*tag- FZ
/*tag- GA
/*tag- GB
/*tag- GC
/*tag- GD
/*tag- GE
/*tag- GF
/*tag- GH
/*tag- GI
/*tag- GJ
/*tag- GK
/*tag- GL
/*tag- GM
/*tag- GN
/*tag- GO
/*tag- GP
/*tag- GQ
/*tag- GR
/*tag- GS
/*tag- GT
/*tag- GU
/*tag- GV
/*tag- GW
/*tag- GX
/*tag- GY
/*tag- GZ
/*tag- HA
/*tag- HB
/*tag- HC
/*tag- HD
/*tag- HE
/*tag- HF
/*tag- HG
/*tag- HH
/*tag- HI
/*tag- HJ
/*tag- HK
/*tag- HL
/*tag- HM
/*tag- HN
/*tag- HO
/*tag- HP
/*tag- HQ
/*tag- HR
/*tag- HS
/*tag- HT
/*tag- HU
/*tag- HV
/*tag- HW
/*tag- HX
/*tag- HY
/*tag- HZ
/*tag- IA
/*tag- IB
/*tag- IC
/*tag- ID
/*tag- IE
/*tag- IF
/*tag- IG
/*tag- IH
/*tag- II
/*tag- IJ
/*tag- IK
/*tag- IL
/*tag- IM
/*tag- IN
/*tag- IO
/*tag- IP
/*tag- IQ
/*tag- IR
/*tag- IS
/*tag- IT
/*tag- IU
/*tag- IV
/*tag- IW
/*tag- IX
/*tag- IY
/*tag- IZ
/*tag- JA
/*tag- JB
/*tag- JC
/*tag- JD
/*tag- JE
/*tag- JF
/*tag- JG
/*tag- JH
/*tag- JI
/*tag- JJ
/*tag- JK
/*tag- JL
/*tag- JM
/*tag- JN
/*tag- JO
/*tag- JP
/*tag- JQ
/*tag- JR
/*tag- JS
/*tag- JT
/*tag- JU
/*tag- JV
/*tag- JW
/*tag- JX
/*tag- JY
/*tag- JZ
/*tag- KA
/*tag- KB
/*tag- KC
/*tag- KD
/*tag- KE
/*tag- KF
/*tag- KG
/*tag- KH
/*tag- KI
/*tag- KJ
/*tag- KK
/*tag- KL
/*tag- KM
/*tag- KN
/*tag- KO
/*tag- KP
/*tag- KQ
/*tag- KR
/*tag- KS
/*tag- KT
/*tag- KU
/*tag- KV
/*tag- KW
/*tag- KX
/*tag- KY
/*tag- KZ
/*tag- LA
/*tag- LB
/*tag- LC
/*tag- LD
/*tag- LE
/*tag- LF
/*tag- LG
/*tag- LH
/*tag- LI
/*tag- LJ
/*tag- LK
/*tag- LL
/*tag- LM
/*tag- LN
/*tag- LO
/*tag- LP
/*tag- LQ
/*tag- LR
/*tag- LS
/*tag- LT
/*tag- LU
/*tag- LV
/*tag- LW
/*tag- LX
/*tag- LY
/*tag- LZ
/*tag- MA
/*tag- MB
/*tag- MC
/*tag- MD
/*tag- ME
/*tag- MF
/*tag- MG
/*tag- MH
/*tag- MI
/*tag- MJ
/*tag- MK
/*tag- ML
/*tag- MM
/*tag- MN
/*tag- MO
/*tag- MP
/*tag- MQ
/*tag- MR
/*tag- MS
/*tag- MT
/*tag- MU
/*tag- MV
/*tag- MW
/*tag- MX
/*tag- MY
/*tag- MZ
/*tag- NA
/*tag- NB
/*tag- NC
/*tag- ND
/*tag- NE
/*tag- NF
/*tag- NG
/*tag- NH
/*tag- NI
/*tag- NJ
/*tag- NK
/*tag- NL
/*tag- NM
/*tag- NO
/*tag- NP
/*tag- NQ
/*tag- NR
/*tag- NS
/*tag- NT
/*tag- NU
/*tag- NV
/*tag- NW
/*tag- NX
/*tag- NY
/*tag- NZ
/*tag- OA
/*tag- OB
/*tag- OC
/*tag- OD
/*tag- OE
/*tag- OF
/*tag- OG
/*tag- OH
/*tag- OI
/*tag- OJ
/*tag- OK
/*tag- OL
/*tag- OM
/*tag- ON
/*tag- OO
/*tag- OP
/*tag- OQ
/*tag- OR
/*tag- OS
/*tag- OT
/*tag- OU
/*tag- OV
/*tag- OW
/*tag- OX
/*tag- OY
/*tag- OZ
/*tag- PA
/*tag- PB
/*tag- PC
/*tag- PD
/*tag- PE
/*tag- PF
/*tag- PG
/*tag- PH
/*tag- PI
/*tag- PJ
/*tag- PK
/*tag- PL
/*tag- PM
/*tag- PN
/*tag- PO
/*tag- PP
/*tag- PQ
/*tag- PR
/*tag- PS
/*tag- PT
/*tag- PU
/*tag- PV
/*tag- PW
/*tag- PX
/*tag- PY
/*tag- PZ
/*tag- QA
/*tag- QB
/*tag- QC
/*tag- QD
/*tag- QE
/*tag- QF
/*tag- QG
/*tag- QH
/*tag- QI
/*tag- QJ
/*tag- QK
/*tag- QL
/*tag- QM
/*tag- QN
/*tag- QO
/*tag- QP
/*tag- QQ
/*tag- QR
/*tag- QS
/*tag- QT
/*tag- QU
/*tag- QV
/*tag- QW
/*tag- QX
/*tag- QY
/*tag- QZ
/*tag- RA
/*tag- RB
/*tag- RC
/*tag- RD
/*tag- RE
/*tag- RF
/*tag- RG
/*tag- RH
/*tag- RI
/*tag- RJ
/*tag- RK
/*tag- RL
/*tag- RM
/*tag- RN
/*tag- RO
/*tag- RP
/*tag- RQ
/*tag- RR
/*tag- RS
/*tag- RT
/*tag- RU
/*tag- RV
/*tag- RW
/*tag- RX
/*tag- RY
/*tag- RZ
/*tag- SA
/*tag- SB
/*tag- SC
/*tag- SD
/*tag- SE
/*tag- SF
/*tag- SG
/*tag- SH
/*tag- SI
/*tag- SJ
/*tag- SK
/*tag- SL
/*tag- SM
/*tag- SN
/*tag- SO
/*tag- SP
/*tag- SQ
/*tag- SR
/*tag- SS
/*tag- ST
/*tag- SU
/*tag- SV
/*tag- SW
/*tag- SX
/*tag- SY
/*tag- SZ
/*tag- TA
/*tag- TB
/*tag- TC
/*tag- TD
/*tag- TE
/*tag- TF
/*tag- TG
/*tag- TH
/*tag- TI
/*tag- TJ
/*tag- TK
/*tag- TL
/*tag- TM
/*tag- TN
/*tag- TO
/*tag- TP
/*tag- TQ
/*tag- TR
/*tag- TS
/*tag- TT
/*tag- TU
/*tag- TV
/*tag- TW
/*tag- TX
/*tag- TY
/*tag- TZ
/*tag- UA
/*tag- UB
/*tag- UC
/*tag- UD
/*tag- UE
/*tag- UF
/*tag- UG
/*tag- UH
/*tag- UI
/*tag- UJ
/*tag- UK
/*tag- UL
/*tag- UM
/*tag- UN
/*tag- UO
/*tag- UP
/*tag- UQ
/*tag- UR
/*tag- US
/*tag- UT
/*tag- UV
/*tag- UW
/*tag- UX
/*tag- UY
/*tag- UZ
/*tag- VA
/*tag- VB
/*tag- VC
/*tag- VD
/*tag- VE
/*tag- VF
/*tag- VG
/*tag- VH
/*tag- VI
/*tag- VJ
/*tag- VK
/*tag- VL
/*tag- VM
/*tag- VN
/*tag- VO
/*tag- VP
/*tag- VQ
/*tag- VR
/*tag- VS
/*tag- VT
/*tag- VU
/*tag- VV
/*tag- VW
/*tag- VX
/*tag- VY
/*tag- VZ
/*tag- WA
/*tag- WB
/*tag- WC
/*tag- WD
/*tag- WE
/*tag- WF
/*tag- WG
/*tag- WH
/*tag- WI
/*tag- WJ
/*tag- WK
/*tag- WL
/*tag- WM
/*tag- WN
/*tag- WO
/*tag- WP
/*tag- WQ
/*tag- WR
/*tag- WS
/*tag- WT
/*tag- WU
/*tag- WV
/*tag- WW
/*tag- WX
/*tag- WY
/*tag- WZ
/*tag- XA
/*tag- XB
/*tag- XC
/*tag- XD
/*tag- XE
/*tag- XF
/*tag- XG
/*tag- XH
/*tag- XI
/*tag- XJ
/*tag- XK
/*tag- XL
/*tag- XM
/*tag- XN
/*tag- XO
/*tag- XP
/*tag- XQ
/*tag- XR
/*tag- XS
/*tag- XT
/*tag- XU
/*tag- XV
/*tag- XW
/*tag- XX
/*tag- XY
/*tag- XZ
/*tag- YA
/*tag- YB
/*tag- YC
/*tag- YD
/*tag- YE
/*tag- YF
/*tag- YG
/*tag- YH
/*tag- YI
/*tag- YJ
/*tag- YK
/*tag- YL
/*tag- YM
/*tag- YN
/*tag- YO
/*tag- YP
/*tag- YQ
/*tag- YR
/*tag- YS
/*tag- YT
/*tag- YU
/*tag- YV
/*tag- YW
/*tag- YX
/*tag- YY
/*tag- YZ
/*tag- ZA
/*tag- ZB
/*tag- ZC
/*tag- ZD
/*tag- ZE
/*tag- ZF
/*tag- ZG
/*tag- ZH
/*tag- ZI
/*tag- ZJ
/*tag- ZK
/*tag- ZL
/*tag- ZM
/*tag- ZN
/*tag- ZO
/*tag- ZP
/*tag- ZQ
/*tag- ZR
/*tag- ZS
/*tag- ZT
/*tag- ZU
/*tag- ZV
/*tag- ZW
/*tag- ZX
/*tag- ZY
/*tag- ZZ
/*tag- AA
/*tag- AB
/*tag- AC
/*tag- AD
/*tag- AE
/*tag- AF
/*tag- AG
/*tag- AH
/*tag- AI
/*tag- AJ
/*tag- AK
/*tag- AL
/*tag- AM
/*tag- AN
/*tag- AO
/*tag- AP
/*tag- AQ
/*tag- AR
/*tag- AS
/*tag- AT
/*tag- AU
/*tag- AV
/*tag- AW
/*tag- AX
/*tag- AY
/*tag- AZ
/*tag- BA
/*tag- BB
/*tag- BC
/*tag- BD
/*tag- BE
/*tag- BF
/*tag- BG
/*tag- BH
/*tag- BI
/*tag- BJ
/*tag- BK
/*tag- BL
/*tag- BM
/*tag- BN
/*tag- BO
/*tag- BP
/*tag- BQ
/*tag- BR
/*tag- BS
/*tag- BT
/*tag- BU
/*tag- BV
/*tag- BW
/*tag- BX
/*tag- BY
/*tag- BZ
/*tag- CA
/*tag- CB
/*tag- CC
/*tag- CD
/*tag- CE
/*tag- CF
/*tag- CG
/*tag- CH
/*tag- CI
/*tag- CJ
/*tag- CK
/*tag- CL
/*tag- CM
/*tag- CN
/*tag- CO
/*tag- CP
/*tag- CQ
/*tag- CR
/*tag- CS
/*tag- CT
/*tag- CU
/*tag- CV
/*tag- CW
/*tag- CX
/*tag- CY
/*tag- CZ
/*tag- DA
/*tag- DB
/*tag- DC
/*tag- DD
/*tag- DE
/*tag- DF
/*tag- DG
/*tag- DH
/*tag- DI
/*tag- DJ
/*tag- DK
/*tag- DL
/*tag- DM
/*tag- DN
/*tag- DO
/*tag- DP
/*tag- DQ
/*tag- DR
/*tag- DS
/*tag- DU
/*tag- DV
/*tag- DW
/*tag- DX
/*tag- DY
/*tag- DZ
/*tag- EA
/*tag- EB
/*tag- EC
/*tag- ED
/*tag- EE
/*tag- EF
/*tag- EG
/*tag- EH
/*tag- EI
/*tag- EJ
/*tag- EK
/*tag- EL
/*tag- EM
/*tag- EN
/*tag- EO
/*tag- EP
/*tag- EQ
/*tag- ER
/*tag- ES
/*tag- ET
/*tag- EU
/*tag- EV
/*tag- EW
/*tag- EX
/*tag- EY
/*tag- EZ
/*tag- FA
/*tag- FB
/*tag- FC
/*tag- FD
/*tag- FE
/*tag- FG
/*tag- FH
/*tag- FI
/*tag- FJ
/*tag- FK
/*tag- FL
/*tag- FM
/*tag- FN
/*tag- FO
/*tag- FP
/*tag- FQ
/*tag- FR
/*tag- FS
/*tag- FT
/*tag- FU
/*tag- FV
/*tag- FW
/*tag- FX
/*tag- FY
/*tag- FZ
/*tag- GA
/*tag- GB
/*tag- GC
/*tag- GD
/*tag- GE
/*tag- GH
/*tag- GI
/*tag- GJ
/*tag- GK
/*tag- GL
/*tag- GM
/*tag- GN
/*tag- GO
/*tag- GP
/*tag- GQ
/*tag- GR
/*tag- GS
/*tag- GT
/*tag- GU
/*tag- GV
/*tag- GW
/*tag- GX
/*tag- GY
/*tag- GZ
/*tag- HA
/*tag- HB
/*tag- HC
/*tag- HD
/*tag- HE
/*tag- HF
/*tag- HG
/*tag- HH
/*tag- HI
/*tag- HJ
/*tag- HK
/*tag- HL
/*tag- HM
/*tag- HN
/*tag- HO
/*tag- HP
/*tag- HQ
/*tag- HR
/*tag- HS
/*tag- HT
/*tag- HU
/*tag- HV
/*tag- HW
/*tag- HX
/*tag- HY
/*tag- HZ
/*tag- IA
/*tag- IB
/*tag- IC
/*tag- ID
/*tag- IE
/*tag- IF
/*tag- IG
/*tag- IH
/*tag- II
/*tag- IJ
/*tag- IK
/*tag- IL
/*tag- IM
/*tag- IN
/*tag- IO
/*tag- IP
/*tag- IQ
/*tag- IR
/*tag- IS
/*tag- IT
/*tag- IU
/*tag- IV
/*tag- IW
/*tag- IX
/*tag- IY
/*tag- IZ
/*tag- JA
/*tag- JB
/*tag- JC
/*tag- JD
/*tag- JE
/*tag- JF
/*tag- JG
/*tag- JH
/*tag- JI
/*tag- JJ
/*tag- JK
/*tag- JL
/*tag- JM
/*tag- JN
/*tag- JO
/*tag- JP
/*tag- JQ
/*tag- JR
/*tag- JS
/*tag- JT
/*tag- JU
/*tag- JV
/*tag- JW
/*tag- JX
/*tag- JY
/*tag- JZ
/*tag- KA
/*tag- KB
/*tag- KC
/*tag- KD
/*tag- KE
/*tag- KF
/*tag- KG
/*tag- KH
/*tag- KI
/*tag- KJ
/*tag- KK
/*tag- KL
/*tag- KM
/*tag- KN
/*tag- KO
/*tag- KP
/*tag- KQ
/*tag- KR
/*tag- KS
/*tag- KT
/*tag- KU
/*tag- KV
/*tag- KW
/*tag- KX
/*tag- KY
/*tag- KZ
/*tag- LA
/*tag- LB
/*tag- LC
/*tag- LD
/*tag- LE
/*tag- LF
/*tag- LG
/*tag- LH
/*tag- LI
/*tag- LJ
/*tag- LK
/*tag- LL
/*tag- LM
/*tag- LN
/*tag- LO
/*tag- LP
/*tag- LQ
/*tag- LR
/*tag- LS
/*tag- LT
/*tag- LU
/*tag- LV
/*tag- LW
/*tag- LX
/*tag- LY
/*tag- LZ
/*tag- MA
/*tag- MB
/*tag- MC
/*tag- MD
/*tag- ME
/*tag- MF
/*tag- MG
/*tag- MH
/*tag- MI
/*tag- MJ
/*tag- MK
/*tag- ML
/*tag- MM
/*tag- MN
/*tag- MO
/*tag- MP
/*tag- MQ
/*tag- MR
/*tag- MS
/*tag- MT
/*tag- MU
/*tag- MV
/*tag- MW
/*tag- MX
/*tag- MY
/*tag- MZ
/*tag- NA
/*tag- NB
/*tag- NC
/*tag- ND
/*tag- NE
/*tag- NF
/*tag- NG
/*tag- NH
/*tag- NI
/*tag- NJ
/*tag- NK
/*tag- NL
/*tag- NM
/*tag- NO
/*tag- NP
/*tag- NQ
/*tag- NR
/*tag- NS
/*tag- NT
/*tag- NU
/*tag- NV
/*tag- NW
/*tag- NX
/*tag- NY
/*tag- NZ
/*tag- OA
/*tag- OB
/*tag- OC
/*tag- OD
/*tag- OE
/*tag- OF
/*tag- OG
/*tag- OH
/*tag- OI
/*tag- OJ
/*tag- OK
/*tag- OL
/*tag- OM
/*tag- ON
/*tag- OO
/*tag- OP
/*tag- OQ
/*tag- OR
/*tag- OS
/*tag- OT
/*tag- OU
/*tag- OV
/*tag- OW
/*tag- OX
/*tag- OY
/*tag- OZ
/*tag- PA
/*tag- PB
/*tag- PC
/*tag- PD
/*tag- PE
/*tag- PF
/*tag- PG
/*tag- PH
/*tag- PI
/*tag- PJ
/*tag- PK
/*tag- PL
/*tag- PM
/*tag- PN
/*tag- PO
/*tag- PP
/*tag- PQ
/*tag- PR
/*tag- PS
/*tag- PT
/*tag- PU
/*tag- PV
/*tag- PW
/*tag- PX
/*tag- PY
/*tag- PZ
/*tag- QA
/*tag- QB
/*tag- QC
/*tag- QD
/*tag- QE
/*tag- QF
/*tag- QG
/*tag- QH
/*tag- QI
/*tag- QJ
/*tag- QK
/*tag- QL
/*tag- QM
/*tag- QN
/*tag- QO
/*tag- QP
/*tag- QQ
/*tag- QR
/*tag- QS
/*tag- QT
/*tag- QU
/*tag- QV
/*tag- QW
/*tag- QX
/*tag- QY
/*tag- QZ
/*tag- RA
/*tag- RB
/*tag- RC
/*tag- RD
/*tag- RE
/*tag- RF
/*tag- RG
/*tag- RH
/*tag- RI
/*tag- RJ
/*tag- RK
/*tag- RL
/*tag- RM
/*tag- RN
/*tag- RO
/*tag- RP
/*tag- RQ
/*tag- RR
/*tag- RS
/*tag- RT
/*tag- RU
/*tag- RV
/*tag- RW
/*tag- RX
/*tag- RY
/*tag- RZ
/*tag- SA
/*tag- SB
/*tag- SC
/*tag- SD
/*tag- SE
/*tag- SF
/*tag- SG
/*tag- SH
/*tag- SI
/*tag- SJ
/*tag- SK
/*tag- SL
/*tag- SM
/*tag- SN
/*tag- SO
/*tag- SP
/*tag- SQ
/*tag- SR
/*tag- SS
/*tag- ST
/*tag- SU
/*tag- SV
/*tag- SW
/*tag- SX
/*tag- SY
/*tag- SZ
/*tag- TA
/*tag- TB
/*tag- TC
/*tag- TD
/*tag- TE
/*tag- TF
/*tag- TG
/*tag- TH
/*tag- TI
/*tag- TJ
/*tag- TK
/*tag- TL
/*tag- TM
/*tag- TN
/*tag- TO
/*tag- TP
/*tag- TQ
/*tag- TR
/*tag- TS
/*tag- TT
/*tag- TU
/*tag- TV
/*tag- TW
/*tag- TX
/*tag- TY
/*tag- TZ
/*tag- UA
/*tag- UB
/*tag- UC
/*tag- UD
/*tag- UE
/*tag- UF
/*tag- UG
/*tag- UH
/*tag- UI
/*tag- UJ
/*tag- UK
/*tag- UL
/*tag- UM
/*tag- UN
/*tag- UO
/*tag- UP
/*tag- UQ
/*tag- UR
/*tag- US
/*tag- UT
/*tag- UV
/*tag- UW
/*tag- UX
/*tag- UY
/*tag- UZ
/*tag- VA
/*tag- VB
/*tag- VC
/*tag- VD
/*tag- VE
/*tag- VF
/*tag- VG
/*tag- VH
/*tag- VI
/*tag- VJ
/*tag- VK
/*tag- VL
/*tag- VM
/*tag- VN
/*tag- VO
/*tag- VP
/*tag- VQ
/*tag- VR
/*tag- VS
/*tag- VT
/*tag- VU
/*tag- VV
/*tag- VW
/*tag- VX
/*tag- VY
/*tag- VZ
/*tag- WA
/*tag- WB
/*tag- WC
/*tag- WD
/*tag- WE
/*tag- WF
/*tag- WG
/*tag- WH
/*tag- WI
/*tag- WJ
/*tag- WK
/*tag- WL
/*tag- WM
/*tag- WN
/*tag- WO
/*tag- WP
/*tag- WQ
/*tag- WR
/*tag- WS
/*tag- WT
/*tag- WU
/*tag- WV
/*tag- WW
/*tag- WX
/*tag- WY
/*tag- WZ
/*tag- XA
/*tag- XB
/*tag- XC
/*tag- XD
/*tag- XE
/*tag- XF
/*tag- XG
/*tag- XH
/*tag- XI
/*tag- XJ
/*tag- XK
/*tag- XL
/*tag- XM
/*tag- XN
/*tag- XO
/*tag- XP
/*tag- XQ
/*tag- XR
/*tag- XS
/*tag- XT
/*tag- XU
/*tag- XV
/*tag- XW
/*tag- XX
/*tag- XY
/*tag- XZ
/*tag- YA
/*tag- YB
/*tag- YC
/*tag- YD
/*tag- YE
/*tag- YF
/*tag- YG
/*tag- YH
/*tag- YI
/*tag- YJ
/*tag- YK
/*tag- YL
/*tag- YM
/*tag- YN
/*tag- YO
/*tag- YP
/*tag- YQ
/*tag- YR
/*tag- YS
/*tag- YT
/*tag- YU
/*tag- YV
/*tag- YW
/*tag- YX
/*tag- YY
/*tag- YZ
/*tag- ZA
/*tag- ZB
/*tag- ZC
/*tag- ZD
/*tag- ZE
/*tag- ZF
/*tag- ZG
/*tag- ZH
/*tag- ZI
/*tag- ZJ
/*tag- ZK
/*tag- ZL
/*tag- ZM
/*tag- ZN
/*tag- ZO
/*tag- ZP
/*tag- ZQ
/*tag- ZR
/*tag- ZS
/*tag- ZT
/*tag- ZU
/*tag- ZV
/*tag- ZW
/*tag- ZX
/*tag- ZY
/*tag- ZZ
/*tag- AA
/*tag- AB
/*tag- AC
/*tag- AD
/*tag- AE
/*tag- AF
/*tag- AG
/*tag- AH
/*tag- AI
/*tag- AJ
/*tag- AK
/*tag- AL
/*tag- AM
/*tag- AN
/*tag- AO
/*tag- AP
/*tag- AQ
/*tag- AR
/*tag- AS
/*tag- AT
/*tag- AU
/*tag- AV
/*tag- AW
/*tag- AX
/*tag- AY
/*tag- AZ
/*tag- BA
/*tag- BB
/*tag- BC
/*tag- BD
/*tag- BE
/*tag- BF
/*tag- BG
/*tag- BH
/*tag- BI
/*tag- BJ
/*tag- BK
/*tag- BL
/*tag- BM
/*tag- BN
/*tag- BO
/*tag- BP
/*tag- BQ
/*tag- BR
/*tag- BS
/*tag- BT
/*tag- BU
/*tag- BV
/*tag- BW
/*tag- BX
/*tag- BY
/*tag- BZ
/*tag- CA
/*tag- CB
/*tag- CC
/*tag- CD
/*tag- CE
/*tag- CF
/*tag- CG
/*tag- CH
/*tag- CI
/*tag- CJ
/*tag- CK
/*tag- CL
/*tag- CM
/*tag- CN
/*tag- CO
/*tag- CP
/*tag- CQ
/*tag- CR
/*tag- CS
/*tag- CT
/*tag- CU
/*tag- CV
/*tag- CW
/*tag- CX
/*tag- CY
/*tag- CZ
/*tag- DA
/*tag- DB
/*tag- DC
/*tag- DD
/*tag- DE
/*tag- DF
/*tag- DG
/*tag- DH
/*tag- DI
/*tag- DJ
/*tag- DK
/*tag- DL
/*tag- DM
/*tag- DN
/*tag- DO
/*tag- DP
/*tag- DQ
/*tag- DR
/*tag- DS
/*tag- DU
/*tag- DV
/*tag- DW
/*tag- DX
/*tag- DY
/*tag- DZ
/*tag- EA
/*tag- EB
/*tag- EC
/*tag- ED
/*tag- EE
/*tag- EF
/*tag- EG
/*tag- EH
/*tag- EI
/*tag- EJ
/*tag- EK
/*tag- EL
/*tag- EM
/*tag- EN
/*tag- EO
/*tag- EP
/*tag- EQ
/*tag- ER
/*tag- ES
/*tag- ET
/*tag- EU
/*tag- EV
/*tag- EW
/*tag- EX
/*tag- EY
/*tag- EZ
/*tag- FA
/*tag- FB
/*tag- FC
/*tag- FD
/*tag- FE
/*tag- FG
/*tag- FH
/*tag- FI
/*tag- FJ
/*tag- FK
/*tag- FL
/*tag- FM
/*tag- FN
/*tag- FO
/*tag- FP
/*tag- FQ
/*tag- FR
/*tag- FS
/*tag- FT
/*tag- FU
/*tag- FV
/*tag- FW
/*tag- FX
/*tag- FY
/*tag- FZ
/*tag- GA
/*tag- GB
/*tag- GC
/*tag- GD
/*tag- GE
/*tag- GH
/*tag- GI
/*tag- GJ
/*tag- GK
/*tag- GL
/*tag- GM
/*tag- GN
/*tag- GO
/*tag- GP
/*tag- GQ
/*tag- GR
/*tag- GS
/*tag- GT
/*tag- GU
/*tag- GV
/*tag- GW
/*tag- GX
/*tag- GY
/*tag- GZ
/*tag- HA
/*tag- HB
/*tag- HC
/*tag- HD
/*tag- HE
/*tag- HF
/*tag- HG
/*tag- HH
/*tag- HI
/*tag- HJ
/*tag- HK
/*tag- HL
/*tag- HM
/*tag- HN
/*tag- HO
/*tag- HP
/*tag- HQ
/*tag- HR
/*tag- HS
/*tag- HT
/*tag- HU
/*tag- HV
/*tag- HW
/*tag- HX
/*tag- HY
/*tag- HZ
/*tag- IA
/*tag- IB
/*tag- IC
/*tag- ID
/*tag- IE
/*tag- IF
/*tag- IG
/*tag- IH
/*tag- II
/*tag- IJ
/*tag- IK
/*tag- IL
/*tag- IM
/*tag- IN
/*tag- IO
/*tag- IP
/*tag- IQ
/*tag- IR
/*tag- IS
/*tag- IT
/*tag- IU
/*tag- IV
/*tag- IW
/*tag- IX
/*tag- IY
/*tag- IZ
/*tag- JA
/*tag- JB
/*tag- JC
/*tag- JD
/*tag- JE
/*tag- JF
/*tag- JG
/*tag- JH
/*tag- JI
/*tag- JJ
/*tag- JK
/*tag- JL
/*tag- JM
/*tag- JN
/*tag-

QY 293 CCGGAGCCGAGGCGAGAGTGCCTCTGTGCGCCGAGCGGAGCCCGGCGACACCCAG 352
 DB 135 ACTGGCCCGGGGGGTGAGATGCCCTTTAAAGCCCTCCGCGGGGAGACCGCCGCC 76
 QY 353 ACGTAGAGCGCGCGCTGCTGCTGACAGTCCGGG 385
 DB 75 GCGCGCCGACCGCCGCGCCGACACACCGCCGG 43
 RESULT 14
 VI0362/c
 ID VI0362 standard; cDNA; 4257 BP.
 AC VI0362:
 DE 30-JUN-1998 (first entry)
 DT Infected cell protein number 4 alpha-4 gene.
 KM Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;
 KW therapeutics; ss.
 OS Herpes simplex virus.
 FH Key Location/Qualifiers
 CDS 361..4257
 /tag= a
 /product= "Infected cell protein"

PN M03804709-A2.
 PD 05-FEB-1998.
 PR 23-JUL-1997; U12904.
 PR 26-JUL-1996; US-690473.
 PA (ARCH-) ARCH DEV CORP.
 PI Leopoldi R, Roizman B;
 DR WPI: 98-130697/12.
 DR P-PSDB: M40200.
 PT Use of herpes simplex virus ICP4 polypeptide - useful for, e.g.
 PR blocking apoptosis in cells, production of proteins and gene therapy
 CC The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein
 CS number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block
 CC cell apoptosis. Similarly the administration of an agent that inhibits
 CC ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.
 CC This can be used for the immortalisation of cells, production of
 CC proteins, gene therapy, or inhibition of cell death induced in vivo.
 CC They can also be used for production of therapeutics comprising
 CC inhibitors of HSV ICP4 function, useful for treating HSV function.
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 13.2%; Score 51.8; DB 1; Length 4257;
 Best Local Similarity 46.5%; Pred. No. 0.19;
 Matches 167; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

DB 27 GCCTTGGGAGCCCTGCGCTGCGCCAGAGCGCGAGCCCGGAGCCCGCTCGGTGAGAC 86
 DB 2806 GCCGCCAGCCCGCCAGGGGGGTGGGGCCCTCGCGGGCGCGGACACAGGCCACGGGGC 2747
 QY 87 TGGGGGTGAGAGTGCCTGCGAGCGTACCCAGCCGCGGAGTACCTCCGCTACACCTGG 146
 DB 2746 GGGGGGGGGCGCGCGCGGGGGCGCGGGGGCGCGGGGGCGGGGGGGGGGGCTGG 2687
 QY 147 GCTGAGTTCCTTGGGTGCGCCGAGAGCGTGGCCGGGCTGAGAGGATGCGGGGGCG 206
 DB 2686 GCCCGGGGGGCTGAGAGGGGGCGGGGGAGGGGGGCGGGGGCTCGAGACGG 2627
 QY 207 GGACGGGGGGGGGGGGGGGGGGCTGTCACTGAGAGAGCGGGGGGGGGGGGGGGGG 266
 DB 2626 GGGCGTCCCGCGCGCTCTTCTTCTTCTTGGGGGGTGGCGGGCGCGCGCTCGGGCGGC 2567
 QY 267 GGGCGCGCCCGCGCTCTTAAAGCGCGAGCGAGCGCGAGGTGCTGTGGCGGC 326
 DB 2566 GGGCGCGCGCGGAGACTTGTGCGCTTGGCGCCCTCCCGGGGGGGGGGGGGGGGG 2507
 QY 327 AGGCGAGGCGCGGGGAGAGCGAGAGCTGAGAGCGCGCGCTGCTGCACTCCGGG 385
 DB 2506 CCGCAGCGCGTGGGGGGGTGGGTGGCTGGCGCGCGCGCGCAGAGGGGGGGCGCAGG 2448

RESULT 15
 VI0362/c
 ID VI0362 standard; DNA; 4257 BP.
 AC VI0362:
 DT 29-JAN-1999 (first entry)
 DE The nucleotide sequence of the Herpes simplex virus ICP4.
 KW Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;
 KW stimulation; inhibition; HSV infection; ss.
 OS Herpes simplex virus.
 FH Key Location/Qualifiers
 CDS 361..4257
 /tag= a
 /product= "HSV ICP4 protein"

PN M03846637-A2.
 PD 22-OCT-1998.
 PR 16-APR-1998; U07573.
 PR 16-APR-1997; US-843659.
 PA (ARCH-) ARCH DEV CORP.
 PI Leopoldi R, Roizman B;
 DR WPI: 98-594559/50.
 DR P-PSDB: W80810.
 PT Use of herpes simplex virus U(s)3 polypeptide - for developing
 PR products for modulating apoptosis in cells and for identifying
 PT compounds which act as stimulators or inhibitors of apoptosis
 PS Example 2; Pages 60-63; 85pp; English.
 CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in
 CC the method of the invention as modulators of apoptosis. The methods
 CC and products can be used to identify compounds which modulate
 CC (stimulate or inhibit) apoptosis in cells. They can be used to
 CC immortalise cells for the study of these cells or for growing cells
 CC in large numbers for the productions of proteins. They can also be
 CC used for stimulating apoptosis in cells, e.g. for treating a subject
 CC with a HSV infection.
 SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T;

Query Match 13.2%; Score 51.8; DB 1; Length 4257;
 Best Local Similarity 46.3%; Pred. No. 0.19;
 Matches 167; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 27 GCCTTGGGAGCCCTGCGCTGCGCCAGAGCGCGAGCCCGGAGCCCGCTCGGTGAGAC 86
 DB 2806 GCCGCCAGCCCGCCAGGGGGGTGGGGCCCTCGCGGGCGCGGACACAGGCCACGGGGC 2747
 QY 87 TGGGGGTGAGAGTGCCTGCGAGCGTACCCAGCCGCGGAGTACCTCCGCTACACCTGG 146
 DB 2746 GGGGGGGGGCGCGCGCGGGGGCGCGGGGGCGCGGGGGCGGGGGGGGGGGCTGG 2687
 QY 147 GCTGAGTTCCTTGGGTGCGCCGAGAGCGTGGCCGGGCTGAGAGGATGCGGGGGCG 206
 DB 2686 GCCCGGGGGGCTGAGAGGGGGCGGGGGAGGGGGGCGGGGGCTCGAGACGG 2627
 QY 207 GGACGGGGGGGGGGGGGGGGGGCTGTCACTGAGAGAGCGGGGGGGGGGGGGGGGG 266
 DB 2626 GGGCGTCCCGCGCGCTCTTCTTCTTCTTGGGGGGTGGCGGGCGCGCGCTCGGGCGGC 2567
 QY 267 GGGCGCGCCCGCGCTCTTAAAGCGCGAGCGAGCGAGGTGCTGTGGCGGC 326
 DB 2566 GGGCGCGCGCGGAGACTTGTGCGCTTGGCGCCCTCCCGGGGGGGGGGGGGGGGG 2507
 QY 327 AGGCGAGGCGCGGGGAGAGCGAGAGCTGAGAGCGCGCGCTGCTGCACTCCGGG 385
 DB 2506 CCGCAGCGCGTGGGGGGGTGGGTGGCTGGCGCGCGCGCGCAGAGGGGGGGCGCAGG 2448

Search completed: September 18, 1999, 00:34:58
 Job time: 1905 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:04; Search time 1962.17 Seconds

(without alignments)
116,740 Million cell updates/sec

Title: US-08-956-518a-102

Sequence: 1 AGCCCTTCCAGCGGGTAG.....CAAGAGTTCCTGCTACATCG 689

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmb1:*
1: gb_dal:*
2: gb_dal2:*
3: gb_om:*
4: gb_om:*
5: gb_pat:*
6: gb_ph:*
7: gb_ph1:*
8: gb_ph2:*
9: gb_ph1:*
10: gb_ph2:*
11: gb_ph3:*
12: gb_ph4:*
13: gb_ph5:*
14: gb_ph6:*
15: gb_ph7:*
16: gb_ph8:*
17: gb_ph9:*
18: gb_ph10:*
19: gb_ph11:*
20: gb_ph12:*
21: gb_ph13:*
22: gb_ph14:*
23: gb_ph15:*
24: gb_ph16:*
25: gb_ph17:*
26: gb_ph18:*
27: gb_ph19:*
28: gb_ph20:*
29: gb_ph21:*
30: gb_ph22:*
31: gb_ph23:*
32: gb_ph24:*
33: gb_ph25:*
34: gb_ph26:*
35: gb_ph27:*
36: gb_ph28:*
37: gb_ph29:*
38: gb_ph30:*
39: gb_ph31:*
40: gb_ph32:*
41: gb_ph33:*
42: gb_ph34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	689	100.0	689	11	AF029839	Homo sapi

2	503	73.0	641	11	AF029838	AF029838 Homo sapi
3	292.6	42.5	137699	11	AC005630	AC005630 Homo sapi
4	282.2	41.0	291	9	HS2E10F	Z56126 H.sapiens C
5	269.4	39.1	192439	35	AC007562	AC007562 Homo sapi
6	261	37.9	178307	35	AC006328	AC006328 Homo sapi
7	237.2	34.4	298	9	HS2G7F	Z60646 H.sapiens C
8	155.4	22.6	1712	11	AF036903	AF036903 Homo sapi
9	154.4	22.4	1836	11	AF037646	AF037646 Homo sapi
10	118.8	17.2	180	9	HS1562F	Z54415 H.sapiens C
11	116.2	16.9	171	9	HS1651F	Z57156 H.sapiens C
12	108	15.7	1559	9	H0M47NAR	L25827 Human a7 n1
13	108	15.7	2087	10	HSAR7A	X70297 H.sapiens m
14	108	15.7	1509	10	HSNACHA7	Y08420 H.sapiens m
15	108	15.7	1876	10	HSU62436	U62436 Human n1cot
16	106.4	15.4	1555	10	HSCHRNA7A	Z23141 H.sapiens C
17	106.4	15.4	1977	10	HSU40583	U40583 Human alpha
18	93	13.5	5281	9	HSU08191	U08191 Human R kap
19	92	13.4	1848	12	M0SNARS	L37663 Mus musculu
20	90.4	13.1	2106	12	RATNARAD	L31619 Rattus rat
21	90.4	13.1	3030	12	S53987	S53987 nicotinic r
22	89.4	13.0	139	9	HS15610R	Z64674 H.sapiens C
23	88.8	12.9	1551	3	BR7	X93604 B.taurus mr
24	74.4	10.8	2696	4	GGA7NAREC	X52295 Chicken alp
25	74.4	10.8	2037	4	GGNNA7A	X68586 G.gallus mr
26	61.6	8.9	2090	4	GGABNAREC	X52296 Chicken alp
27	60.4	8.8	189	9	HS151DR	Z61392 H.sapiens C
28	58.2	8.4	3029	37	AF143847	AF143847 Helicobis
29	50	7.3	89871	37	AC005890	AC005890 Drosophi
30	48.2	7.0	32039	1	SC7A1	AL034447 Streptomy
31	46.4	6.7	336	14	G47584	G47584 Z25357_1 Ze
32	46.2	6.7	76947	36	AC004326	AC004326 Drosophi
33	46.2	6.7	3629	37	AF143846	AF143846 Helicobis
34	44.6	6.5	2493	12	RNNIACERE	X15834 R.norvegicu
35	44.6	6.5	2461	12	RNNI2976	U42976 Rattus norv
36	44	6.4	1008	12	RATNACHRRS	M33952 Rat neutroa
37	42.8	6.2	143664	35	AC007291	AC007291 Drosophi
38	42.6	6.2	1574	11	AF069682	AF069682 Homo sapi
39	42.4	6.2	3459	10	HS012582	AF012582 Homo sapi
40	42.4	6.2	785	10	HS013727	A113727 Homo sapi
41	42.2	6.1	107889	11	AC004519	AC004519 Homo sapi
42	42.2	6.1	42301	11	AC005943	AC005943 Homo sapi
43	41.6	6.0	120766	42	AC004150	AC004150 Homo sapi
44	41.2	6.0	2672	3	RAT7HBP	J05602 Rabbit mult
45	41.2	6.0	4085	4	CHKACHR02	K02904 Chicken nic

ALIGNMENTS

RESULT 1
AF029839 689 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION AF029839
ACCESSION AF029839
NID 93757794
VERSION AF029839.1 GI:3757794
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
1 (bases 1 to 689)
Moore,T., Jacobs,S., Merlweher,J., Choi,M.J., Kim,E.J., Walton,K.,
Butting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
TITLE Genomic organization and partial duplication of the human alpha7
neural nicotinic acetylcholine receptor gene
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
REFERENCE 2 (bases 1 to 689)
AUTHORS Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
Breese,C., Davis,A., Hopkins,J. and Freedman,R.
TITLE Direct Submission


```

OY      536  -----TTCCATTGCAATACAGCATTTGGTGGATAGCTGGAACTGGATATTCGTGA 588
          |||||||
DB      481  TTTTAGATTGCATTTGCTATCAGCATTTTGATGATAGCTGGAAACTGGATATTCGTGA 540
          |||||||
OY      589  TGAGCGCTTTGACGCCACATTCACACTAACGTGTGGTGAATTCCTTCCGGGCATTGGCCA 648
          |||||||
DB      541  TGAGGCGTTTGACGCCACATTCACACTAACGTGTGGTGAATTCCTTCCGGGCATTGGCCA 600
          |||||||
DB      601  GTACCTGCCTCCAGGCATATTCACAAGATTCCTGCTACATCG 641
          |||||||

RESULT   3
LOCUS    AC005630                                DNA                PRI        15-JAN-1999
DEFINITION Homo sapiens clone Dll129D05, complete sequence.
SESSION  AC005630
          G4159882
VERSION  AC005630.1 GI:4159882
KEYWORDS HTG.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 137699)
AUTHORS  Waterston,R.H.
TITLE    The sequence of Homo sapiens clone
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 137699)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (08-SEP-1998) Genome Sequencing Center, Washington
          University School of Medicine, 444 Forest Park Parkway, St. Louis
          MO 63108, USA
REFERENCE 3 (bases 1 to 137699)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (15-JAN-1999) Genome Sequencing Center, Washington
          University School of Medicine, 444 Forest Park Parkway, St. Louis
          MO 63108, USA
COMMENT  On Jan 15, 1999 this sequence version replaced gi:3550965.
FEATURES             Location/Qualifiers
     source           1..137699
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Dll129D05"
BASE COUNT  36908 a 30670 c 32412 g 37709 t
ORIGIN

```

Query Match	42.5%	Score 292.6	DB 11	Length 137699
Best Local Similarity	83.1%	Prod. No. 6.5e-612		
Matches 345	Conservative 0	Mismatches 69	Indels 1	Gaps 1
QY 1	AGCCCTTCCACGAGGGGTAGCGGGGACAGTGTGCTGTGCCCTTTAAACTCGGGCTTG	60		
Db 47346	AGCCCTTTCACGAGCGATAGCGGGGCTGTGTGCTGTGCCCTTTAAAGTGGGGCTTG	47405		
QY 61	ACGGAGCGCGGCGCTTCGTGCTGTGAGTCCGGTTATAAGGAGCAGCCCGCAGGGCCGC	120		
Db 47406	ACAGGAGCAGCGCTTCCTCTGTGTGAGTCTTTAAAGGAGCAGCGCCGACAGGCCGC	47465		
QY 121	CACATTAAGCT-CCGCGCAAGTCTTCGGTGGCCCTTGCCATTTCACAGCCGCTCCACAGA	179		
Db 47466	CACACAGCTCCCGCAGAGGCGCTCGGTGCCCTTGCCATTTCACAGCCCTAATCCGAGTA	47525		
QY 189	GGGTACAGGGCGGGGAGAGGTGAGACCGCAGAGACTGTGGCCGGGGGGCCCGGCTGGTG	239		
Db 47526	GAGTTTGGGCATCAGGAGAAAGCGGAGGCTGGGAGAGACGGCCGCGAAGGTTCCTCCGCGGGT	47585		
QY 240	GCGGCGGCATGACAGCGGCTCGGCACTGGTCTCTTTCCGCGGCCCTCCGCGGAGGT	299		

Db	47586	gattggccgcggtacacagcgcgtccgacgsggctcacacttccggcgccctcccgccagaggt	47645
Qy	300	GAGGGGAAGATGTCCATGTCAAGGTTCAAGGCCAAACCGAAGTACTGGCCCTCTATCTTC	359
Db	47646	GAGATTAATATGTCGTGTGAGGTTCAAGGCCAAGCTGAGAGTTGTGGCCCTCATCTTC	47705
Qy	360	CAGGAACACAGAGCCACAGCGCGGCTCAAGCCCAACCCGAACATTAAATTA	414
Db	47706	CACAAAGACAGAGACCCCGCCCGCAGCTCAAGCTTCACGTGCAATCATCAGGTGA	47760

RESULT	4
HS2E10F	
LOCUS	HS2E10F 291 bp DNA PRI 18-OCT-1995
DEFINITION	H.sapiens Cpg island DNA genomic MseI fragment, clone ze10, forward read cpgze10.f.tild.
ACCESSION	Z58126
NID	G1029357
VERSION	Z58126.1 GI:1029357
KEYWORDS	Cpg Island; genomic MseI fragment.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 291)
JOURNAL	MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
REFERENCE	Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS	CB10 1RQ, England. E-mail contact: humphreys@sanger.ac.uk
TITLE	2 (bases 1 to 291)
JOURNAL	Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
REFERENCE	Purification of CpG islands using a methylated DNA binding column
AUTHORS	Nat. Genet. 6 (3), 236-244 (1994)
TITLE	9428270
JOURNAL	vector: pGEM-5zf(-)
COMMENT	

Clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1HQ, UK. See URL: <http://www.hgmp.mrc.ac.uk/> for details or contact: blonhelp@hgmp.mrc.ac.uk.

FEATURES	Location/Qualifiers
source	1. .291

BASE COUNT	ORIGIN
44 a	95 c 103 g 45 t 4 others

Query Match	41.0%	Score 282.2;	DB 9;	Length 291;
Best Local Similarity	97.6%;	Pred. No. 1.7e-58;		
Matches 284; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0

QY	47	TAACTGGGCTTACGAGAGCCGGCTCTCTGTGGATGATCGTTAAAGAGAGCA	106
	1	TAACTGGGCTTACGAGAGCCGGCTCTCTGTGGATGATCGTTAAAGAGAGCA	60
QY	107	GGCCGCGAGGCGGCACATAGCTCCGCGCAAGTCTGTGGTGGCCCTGTCATTTCCAGC	166
Db	61	GNCCGCGAGGCGGCACATAGCTCCGCGCAAGTCTGTGGTGGCCCTGTCATTTCCAGC	120
QY	167	CGCGCTCCACAGAGGATCACGGCGCGGAGAGTGTAGACCGCGAGAGCTCGGCGGGG	226
Db	121	CGCGCTCCACAGAGGATCACGGCGCGGAGAGTGTAGACCGCGAGAGCTCGGCGGGG	180
QY	227	GCCCGGCTGTGGCGCGGCCAATGACAGGCGCTCGGGAACTGTGCTCTTTCCGCGCCC	286
Db	181	GNCCGCGCTGTGTGTGGCGGCAATGACAGGCGCTCGGGAAAGGCTCTTTTCCGCGCCC	240

Db 181 GNCCGCGTGGTGC GGCNCATGACACGCGGCTCGGGACAGGCTCCTTTTCGCGGCCCC 240


```
OY 287 TCCCGCGGAGGTGAGGGAAGATGTCATGTCAGGGCTTCAAGGCCAAC 337
|||||
Db 241 TCCCGCGGAGGTGAGGGAAGATGTCATGTCAGGGCTTCAAGGCCAAC 291

RESULT 5
AC007562
LOCUS AC007562 192439 bp DNA HTG 05-JUN-1999
DEFINITION Homo sapiens clone NH0497C14, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
AC007562
NID 95001485
VERSION AC007562.2 GI:5001485
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 192439)
Waterston, R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 192439)
Waterston, R.H.
Unpublished
Submitted (14-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 5, 1999 this sequence version replaced g1:4827309.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 5335: contig of 5335 bp in length
* 5336 5353: gap of unknown length
* 5354 20005: contig of 14652 bp in length
* 20006 20023: gap of unknown length
* 20024 98080: contig of 78057 bp in length
* 98081 98098: gap of unknown length
* 98099 192439: contig of 94541 bp in length.
FEATURES
source
1..192439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0497C14"
BASE COUNT 55693 a 40710 c 39187 g 56795 t 54 others
ORIGIN

Query Match 39.1%; Score 269.4; DB 35; Length 192439;
Best Local Similarity 81.2%; Pred. No. 2.5e-55;
Matches 337; Conservative 0; Mismatches 76; Indels 2; Gaps 2;

OY 1 AGCCCTTCCAGCGGTAGCGGGGAGTGTGTCGTCCTTTAACTGCGGCTTG 60
|||||
Db 30173 AGCCCTTCCAGCGGTAGCGGGGAGTGTGTCGTCCTTTAACTGCGGCTTG 30232
|||||
OY 61 AGCGGAGCGGCGCTCTGTCGTGAGTGTGTTAAAGGAGACAGCCCGAGGCGCG 120
|||||
Db 30233 ACAGGAGCAGCGGCTCTGTCGTGAGTGTGTTAAAGGAGACAGCCCGAGGCGCTC 30292
|||||
OY 121 CACATAGCT-CCGCGCAAGCTCGGTGCGCCCTTGCCATTTCCAGCGCGGTCCACAGA 179
|||||
Db 30293 CACACAGCTCCCGCAGAGAGCGCTGTGTCGTCCTTTCCAGCGCTTACTGTGACTA 30332
|||||
OY 180 GGGTCACGGCGGCGGAGAGTGAAGCGGACAGAGCTGCGCGGGGCGCCCGCTGTG 239
|||||
Db 30353 GAGTTGAGCGAGAGAGAGGAGAGTGGAGAGACACCGCGGAGAGAGTCCCGGGGGTG 30412
|||||
```

```
OY 240 GCCGGGCGATGACAGCGGCTCGGAGACTGCTCTTCCGCGCCCTCCCGGAGAGT 299
|||||
Db 30413 GTTGGGCGGTGACAGCGGCTCGGAGAGTGTGCTCTTCCGCGCCCTCCCGAGAGT 30471
|||||
OY 300 GAGGGAAGATGTCATGTCAGGTTCAAGGCCAAACCAAGTACTGCGCTCATCTTC 359
|||||
Db 30472 GAGATTAATGTCGTCGTGAGGTTCAAGGCCAAAGCTGAAGTGTGATCTATCTTC 30331
|||||
OY 360 CAGGAGAACAGGAGCCACAGCGGCTCAGCGCCCAACCAATTAGATTAA 414
|||||
Db 30532 CACAAGAACAGGAGCGGCTCGGAGAGTGTGCTCTTCCAGCTCAACATCAGCTGA 30586
|||||

RESULT 6
AC006328
LOCUS AC006328 178307 bp DNA HTG 13-MAR-1999
DEFINITION Homo sapiens clone NH0102005, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
AC006328
NID 94415869
VERSION AC006328.2 GI:4415969
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 178307)
Waterston, R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 178307)
Waterston, R.H.
Unpublished
Submitted (09-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 1999 this sequence version replaced g1:4138780.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1163: contig of 1163 bp in length
* 1164 1181: gap of unknown length
* 1182 57065: contig of 45885 bp in length
* 57067 57084: gap of unknown length
* 57085 178307: contig of 121223 bp in length.
FEATURES
source
1..178307
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0102005"
BASE COUNT 49986 a 38822 c 39355 g 50108 t 36 others
ORIGIN

Query Match 37.9%; Score 261; DB 35; Length 178307;
Best Local Similarity 80.0%; Pred. No. 2.7e-53;
Matches 332; Conservative 0; Mismatches 80; Indels 3; Gaps 2;

OY 1 AGCCCTTCCAGCGGTAGCGGGGAGTGTGTCGTCCTTTAACTGCGGCTTG 60
|||||
Db 10173 AGCCCTTCCAGCGGTAGCGGGGAGTGTGTCGTCCTTTAACTGCGGCTTG 10232
|||||
OY 61 AGCGGAGCGGCGCTCTGTCGTGAGTGTGTTAAAGGAGACAGCCCGAGGCGCG 120
|||||
Db 10233 ACAGGAGCAGCGGCTCTGTCGTGAGTGTGTTAAAGGAGACAGCCCGAGGCGCTC 10292
|||||
OY 121 CACATAGCT-CCGCGCAAGCTCGGTGCGCCCTTGCCATTTCCAGCGCGCTCCACAGA 179
|||||
```



```

Db 10293 CACACAGTCCCCGAGAGGCGCTTGTCCTTGCCATTTCACAGCCCTACTGTGACTA 10352
Qy 180 GGGTCACGGCGGCGGAGAGTGGAGCCCGAGAGCTTCGCCGGGGGCGCCCTGTG 239
Db 10353 GAGTTAGGACACAGAGAGGCGGAGCTGGGAGACACCCGAGAGTCCCGCGGGTG 10412
Qy 240 GCGGGGGCCATACAGCGGCTGGGACTGGCTCTTTCCGGCCCCCTCCCGCGAGGT 299
Db 10413 GTTGGCGCCGTACAGCGGCTCCG--ACGGGCTCACCTTCGCCGCCACCTCCAGAGGT 10470
Qy 300 GAGGGAGAGATGTCATGTCATGTTCAAGGCCCAACCAAGTACTGGCTCTATCTTC 359
Db 10471 GAGATTAATATCCGTGTGAGGTTCAAGGCCCAAGTACTGTGTGAGCTCTATCTTC 10530
Qy 360 CAGGAGAACGAGGACACAGCGCGGCTCACGCCCAACCCGACATTAAGATTA 414
Db 10531 CACAAGAACGAGAGCGCTGCCGAGCTCATGTCCACTGCACATCATCAGTTGA 10585

LOCUS       HS2G7F      298 bp      DNA      PRI      19-OCT-1995
DEFINITION  H.sapiens Cpg Island DNA genomic MseI fragment, clone 2g7, forward
ACCESSION   Z60646
VERSION     Z60646.1 GI:1032750
KEYWORDS    Cpg island; genomic MseI fragment.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 298)
AUTHORS     MacDonald,M., Huckle,E., Wilkinson,P. and Mcklem,G.
TITLE       Direct Submission
JOURNAL     Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
2 (bases 1 to 298)
AUTHORS     Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE       Purification of Cpg Islands using a methylated DNA binding column
JOURNAL     Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE     94282070
COMMENT     Vector: pGEM-52f(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hmp.mrc.ac.uk/ for details
            or contact: biolhelp@hmp.mrc.ac.uk.
FEATURES             location/Qualifiers
             source          1..298
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /sex="male"
                        /dev_stage="adult"
                        /tissue_type="blood"
                        /clone_lib="CGI-1"
                        /clone="2g7"
BASE COUNT      47 a      99 c      102 g      48 t      2 others
ORIGIN
Query Match      34.4%; Score 237.2; DB 9; Length 298;
Best Local Similarity 94.7%; Pred. No. 1.2e-47;
Matches 288; Conservative 0; Mismatches 10; Indels 6; Gaps 4;
Qy 48 AACGCGGCTGAGGAGCGCGGCGCTCCCTGCGTGAGGAGGATTAAGGAGAGAG 107
Db 1 AATGCGGCTGAGGAGCGCGGCGG-NCTTCCTGTGAGTGTGATTAAGGAGAGCG 59
Qy 108 CCGCGGAGGCGGACATAGCTCCGCGCAAGTCTGGTGGCCCTTGCCATTTCCAGCC 167
Db 60 CCGCGCA-GCGCGCACATAGCTCCGCGCAAGTCTGGTGGCCCTTGCCATTTCCAGCC 118
Qy 168 GCGTCCACGAGGTACGCGCGCGGAGAGGTGAGACCGGAGAGCTCGCGGAGG 227

```

```

Db 119 GCGTCCACGAGGCTACGCGGCGGAGAGGTGGAGCCGCGAGANTCGCC--GGG 175
Qy 228 CCGGCGCTGTGGCGGCGGCGCATGACAGCGGCTGCGGAGCTGCTCTTTCCGCCCT 287
Db 176 GCGGCGCTGTGGTGGCGGCGCATGACAGCGGCTGCGGAGAGCTCTTTCCGCCCT 235
Qy 288 CCGGCGGAGGTGAGGAGAGATGTCATGTGTCAGGTTCAAGGCCAAGCGAGTACTG 347
Db 236 CCGGCGGCA-GTGAAGGAGAGATGTCCATGTCCGGGTTCAAGGCCAAGCGAGTACTG 294
Qy 348 GCGT 351
Db 295 GCGT 298

RESULT      8
LOCUS       AF036903      1712 bp      mRNA      PRI      16-OCT-1998
DEFINITION  Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
ACCESSION   AF036903
VERSION     93757807
KEYWORDS    AF036903.1 GI:3757807
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1712)
AUTHORS     Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
            Moore,T., Jacobs,S., Meliether,J., Choi,M.J., Kim,E.J., Walton,K.,
            Bulfinch,R., Davis,A., Breese,C., Freedman,R. and Leonard,S.
TITLE       Genomic organization and partial duplication of the human alpha7
            neuronal nicotinic acetylcholine receptor gene (CHRNA7)
JOURNAL     Genomics 52 (2), 173-185 (1998)
MEDLINE     99000837
REFERENCE   2 (bases 1 to 1712)
AUTHORS     Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
            Breese,C., Davis,A., Hopkins,J. and Freedman,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
            Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
FEATURES             location/Qualifiers
             source          1..1712
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /db_xref="dbEST:M52861"
                        /db_xref="dbEST:AA037389"
                        /chromosome="15"
                        /clone="321242"
                        /clone_lib="Soares papathroid tumor NbhPA library,
                        Reseach Genetics/IMAGE Consortium, LLNL"
                        <1..1712
                        /note="alpha-7 neuronal nicotinic acetylcholine receptor;
                        alternatively spliced"
BASE COUNT      356 a      489 c      457 g      410 t
ORIGIN
Query Match      22.6%; Score 155.4; DB 11; Length 1712;
Best Local Similarity 99.4%; Pred. No. 6.3e-28;
Matches 156; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 533 CAATTCATATGCTAATCCACAGATTTGTGATTAAGTGAACAGTCCGATATGCTGATGAG 592
Db 45 CAGTTCATATGCTAATCCACAGATTTGTGATTAAGTGAACAGTCCGATATGCTGATGAG 104
Qy 593 GCGTTGAGCGCACATTCACACATTAAGTGTGATTTGATTTCTTGAGGCAATGCGAGTAC 652
Db 105 GCGTTGAGCGCACATTCACACATTAAGTGTGATTTGATTTCTTGAGGCAATGCGAGTAC 164
Qy 653 CTGCTCCAGGACATATTCAGAGTTCCTGCTACATCG 689

```



```

DB 165 CTGGCTCCAGCATATTCAGAGTCTCTCATCG 201
|||||
RESULT 9
AF037646 1836 bp RNA PRI 16-OCT-1998
LOCUS AF037646 Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION precursor RNA, partial sequence.
ACCESSION AF037646
NID AF037646.1 GI:3757808
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
1 (bases 1 to 1836)
Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene (CHRNA7)
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
FEATURES
SOURCE 2 (bases 1 to 1836)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
LOCATION/Qualifiers
1. 1836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:M03952"
/db_xref="dbEST:N73891"
/chromosome="15"
/clone="297134"
/clone_1b="Soares fetal liver spleen library, Research
Genetics/IMAGE Consortium, LBNL"
/note="unprocessed mRNA with intron"
<1. 1836
/note="alpha-7 neuronal nicotinic acetylcholine receptor
precursor; intron included in 5' region"
misc_feature
<1. 1836
BASE COUNT 383 a 503 c 469 g 481 t
ORIGIN
Query Match 22.4%; Score 154.4; DB 11; Length 1836;
Best Local Similarity 99.4%; Pred. No. 1.1e-27;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

VERSION 254415.1 GI:1020456
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS MacDonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
1 (bases 1 to 180)
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquyer@sanger.ac.uk
REFERENCE
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
2 (bases 1 to 180)
Purification of Cpg islands using a methylated DNA binding column
NAT. GENET. 6 (3), 236-244 (1994)
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.
LOCATION/Qualifiers
1. 180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_1b="CGI-1"
/clone="11592"
BASE COUNT 33 a 57 c 58 g 29 t 3 others
ORIGIN
Query Match 17.2%; Score 118.8; DB 9; Length 180;
Best Local Similarity 87.0%; Pred. No. 3.8e-19;
Matches 140; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

```

```

DB 165 CTGGCTCCAGCATATTCAGAGTCTCTCATCG 201
|||||
RESULT 9
AF037646 1836 bp RNA PRI 16-OCT-1998
LOCUS AF037646 Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
DEFINITION precursor RNA, partial sequence.
ACCESSION AF037646
NID AF037646.1 GI:3757808
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
1 (bases 1 to 1836)
Moore, T., Jacobs, S., Meriwether, J., Choi, M., Kim, E., Walton, K.,
Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene (CHRNA7)
JOURNAL Genomics 52 (2), 173-185 (1998)
MEDLINE 99000837
FEATURES
SOURCE 2 (bases 1 to 1836)
Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
Breese, C., Davis, A., Hopkins, J. and Freedman, R.
Direct Submission
Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
LOCATION/Qualifiers
1. 1836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="dbEST:M03952"
/db_xref="dbEST:N73891"
/chromosome="15"
/clone="297134"
/clone_1b="Soares fetal liver spleen library, Research
Genetics/IMAGE Consortium, LBNL"
/note="unprocessed mRNA with intron"
<1. 1836
/note="alpha-7 neuronal nicotinic acetylcholine receptor
precursor; intron included in 5' region"
misc_feature
<1. 1836
BASE COUNT 383 a 503 c 469 g 481 t
ORIGIN
Query Match 22.4%; Score 154.4; DB 11; Length 1836;
Best Local Similarity 99.4%; Pred. No. 1.1e-27;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

RESULT 11
HS165HIF 171 bp DNA PRI 18-OCT-1995
LOCUS HS165HIF H.sapiens Cpg island DNA genomic MseI fragment, clone 165h1,
DEFINITION forward read cpj165h1.ftla.
ACCESSION 257156
NID 91028387
VERSION 257156.1 GI:1028387
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.
1 (bases 1 to 171)
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquyer@sanger.ac.uk
REFERENCE
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
2 (bases 1 to 171)
Purification of Cpg islands using a methylated DNA binding column
NAT. GENET. 6 (3), 236-244 (1994)

```


MEDLINE

94282070
Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See URL: <http://www.hgmp.mrc.ac.uk/> for details or contact: blhelp@hgmp.mrc.ac.uk.

FEATURES

Source
Location/Qualifiers
1. 171
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_id="CGI-1"
/clone="165h1"
BASE COUNT 30 a 56 c 57 g 27 t 1 others
ORIGIN

Query Match 16.9%; Score 116.2; DB 9; Length 171;
Best Local Similarity 87.3%; Pred. No. 1.6e-18;
Matches 138; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

47 TAACTGGGCTTACCGGAGCGCCCTCTCTGTCGGTGGAGTGGTTATTAAGGAGCA 106
|||||
1 TAACTGGGCTTACAGGGGCGCGCTCTCTGTCGGTGGAGTGGCTTAAAGGAGCA 60

107 GCCCGGAGCGCGCCACATAGCTCCCGCAAGTCCGCTGCCCTTGCCATTTTCCAGC 166
|||||
61 GNCGCCAGCGCGCCACACAGACTCCCGCAAGACTT-GGTGCCCTTGCCATTTTCCAGC 119

167 CGCGCTCCACAGAGGTCACGCGCGCGGAGAGAGTGG 204
|||||
120 CGCGCTCCACAGAGTGAAGGCTGACGAGAGAGCGG 157

RESULT 12

HUMAN7NAR 1559 bp mRNA PRI 18-AUG-1994
LOCUS Human a7 nicotinic acetylcholine receptor mRNA.
DEFINITION L25827
ACCESSION 9438616
VERSION L25827.1 GI:438616
KEYWORDS nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.

7CE Homo sapiens (library: Clontech HL1065b; ATCC 37433) fetus brain CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (sites)
Dovelette-Stamm, L., Monteggia, L.M., Donnelly-Roberts, D., Wang, M.T., Lee, J., Tian, J. and Giordano, T.
Cloning and sequence of the human a7 nicotinic acetylcholine receptor
Res. (1993) In press
Location/Qualifiers
1. 1559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetus"
/tissue_type="brain"
/tissue_id="Clontech HL1065b; ATCC 37433"

JOURNAL Drug Dev. Res. (1993) In press
FEATURES
Source
Location/Qualifiers
1. 1559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetus"
/tissue_type="brain"
/tissue_id="Clontech HL1065b; ATCC 37433"

BASE COUNT 314 a 463 c 440 g 342 t
ORIGIN

Query Match 15.7%; Score 108; DB 9; Length 1559;
Best Local Similarity 95.7%; Pred. No. 1.6e-16;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
574 CTGCGATTGCTGATGAGCGCTTGACGCCACATTCACACTAAGCTGTGTGAATTC 633
|||||

DB 319 CTATTAACAGTGTCTGATGAGCGCTTTGACGCCACATTCACACTAAGCTGTGTGAATTC 378
QY 634 TTCTGGCATTTGCCAGTACCTGCTCCAGGCATATTCAGAGTTCCTGTACATCG 689
|||||
DB 379 TTCTGGCATTTGCCAGTACCTGCTCCAGGCATATTCAGAGTTCCTGTACATCG 434

RESULT 13

HSARA7A 2087 bp mRNA PRI 01-JUN-1994
LOCUS H.sapiens mRNA for neuronal nicotinic acetylcholine receptor
DEFINITION alpha-7 subunit.
ACCESSION X70297
NID 9496606
VERSION X70297.1 GI:496606
KEYWORDS neuronal nicotinic acetylcholine receptor alpha-7 subunit.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 2087)
Peng, X., Katz, M., Gerzanich, V., Anand, R. and Lindstrom, J.
Human alpha 7 acetylcholine receptor: cloning of the alpha 7 subunit from the SH-SY5Y cell line and determination of pharmacological properties of native receptors and functional alpha 7 homomers expressed in Xenopus oocytes
Mol. Pharmacol. 45 (3), 546-554 (1994)
94195283
2 (bases 1 to 2087)
Katz, M.
Direct Submission
Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235 Stemmler Hall, Philadelphia, PA 19104, USA
Location/Qualifiers
1. 2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/cell_line="SH-SY-5Y"
/clone_id="lambda zap II"
/clone="SHSY3"
104. 175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit.
104. 1612
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit."
/protein_id="CAA49778.1"
/db_xref="PIR:9496607"
/db_xref="PDB:9496607"
/db_xref="GI:496607"
/translation="MRCSPGVWGLGAASLHVSLGGEFORLKELYKNVPLRPV
ANDSPLTYFSLIOLIMDVDEKNVLTITNIMLSMTDHYLQNVSEPVKTVRF
PDGQIMKPDILLYNSADRFDTFRNTLVNNSGHCQYLPRTFSSCITDVRMPFD
VOCHKLFGSNTSGMSLDLQMDLQDISITNGSMDVGLGKRSERYECCCKPYP
DVTFTVTKRRRLTYGLNLIPCVLISALALVFLPADSGKISLIGITVLISLTFV
LLVAETMPATSDSVPLIAOYFASMTIYGLSVVTVIVLYQHHHPDGGKMPKTRVI
LNMCAWFLMKRPGEDEVKVPACQKORCSLSAVEMSAVGGPPNSGMLTYGRGL
DGVHCVPTPDGSGVYCGRAKSPTHDEHLLHGQPEGGPDIAKILEVRYIANRRRCQ
DSEAVSCSMKRAACVYDRCLMAFVSFTIITIGLIMSAPFVZAVSKDPA"

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
94195283
2 (bases 1 to 2087)
Katz, M.
Direct Submission
Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of Medicine, Dept of Neuroscience, 36th & Hamilton Walk, Rm 235 Stemmler Hall, Philadelphia, PA 19104, USA
Location/Qualifiers
1. 2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/cell_line="SH-SY-5Y"
/clone_id="lambda zap II"
/clone="SHSY3"
104. 175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit.
104. 1612
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit."
/protein_id="CAA49778.1"
/db_xref="PIR:9496607"
/db_xref="PDB:9496607"
/db_xref="GI:496607"
/translation="MRCSPGVWGLGAASLHVSLGGEFORLKELYKNVPLRPV
ANDSPLTYFSLIOLIMDVDEKNVLTITNIMLSMTDHYLQNVSEPVKTVRF
PDGQIMKPDILLYNSADRFDTFRNTLVNNSGHCQYLPRTFSSCITDVRMPFD
VOCHKLFGSNTSGMSLDLQMDLQDISITNGSMDVGLGKRSERYECCCKPYP
DVTFTVTKRRRLTYGLNLIPCVLISALALVFLPADSGKISLIGITVLISLTFV
LLVAETMPATSDSVPLIAOYFASMTIYGLSVVTVIVLYQHHHPDGGKMPKTRVI
LNMCAWFLMKRPGEDEVKVPACQKORCSLSAVEMSAVGGPPNSGMLTYGRGL
DGVHCVPTPDGSGVYCGRAKSPTHDEHLLHGQPEGGPDIAKILEVRYIANRRRCQ
DSEAVSCSMKRAACVYDRCLMAFVSFTIITIGLIMSAPFVZAVSKDPA"

FEATURES

Source
Location/Qualifiers
1. 2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/cell_line="SH-SY-5Y"
/clone_id="lambda zap II"
/clone="SHSY3"
104. 175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit.
104. 1612
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit."
/protein_id="CAA49778.1"
/db_xref="PIR:9496607"
/db_xref="PDB:9496607"
/db_xref="GI:496607"
/translation="MRCSPGVWGLGAASLHVSLGGEFORLKELYKNVPLRPV
ANDSPLTYFSLIOLIMDVDEKNVLTITNIMLSMTDHYLQNVSEPVKTVRF
PDGQIMKPDILLYNSADRFDTFRNTLVNNSGHCQYLPRTFSSCITDVRMPFD
VOCHKLFGSNTSGMSLDLQMDLQDISITNGSMDVGLGKRSERYECCCKPYP
DVTFTVTKRRRLTYGLNLIPCVLISALALVFLPADSGKISLIGITVLISLTFV
LLVAETMPATSDSVPLIAOYFASMTIYGLSVVTVIVLYQHHHPDGGKMPKTRVI
LNMCAWFLMKRPGEDEVKVPACQKORCSLSAVEMSAVGGPPNSGMLTYGRGL
DGVHCVPTPDGSGVYCGRAKSPTHDEHLLHGQPEGGPDIAKILEVRYIANRRRCQ
DSEAVSCSMKRAACVYDRCLMAFVSFTIITIGLIMSAPFVZAVSKDPA"

BASE COUNT 441 a 601 c 573 g 472 t
ORIGIN

Query Match 15.7%; Score 108; DB 10; Length 2087;
Best Local Similarity 95.7%; Pred. No. 1.6e-16;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
574 CTGCGATTGCTGATGAGCGCTTGACGCCACATTCACACTAAGCTGTGTGAATTC 633
|||||

Db 445 CTATACAGTCTGATGAGCCCTTTGACGCCATTCACACTACAGCTGTGTAATTC 504
OY 634 TTCTGGGCAATGGCAGTACCTGCGCCAGCATATTCAGAGTCTCGCTACATCG 689
Db 505 TTCTGGGCAATGGCAGTACCTGCGCCAGCATATTCAGAGTCTCGCTACATCG 560

RESULT 14
LOCUS HSNACHRA7 1509 bp mRNA PRI 22-JAN-1998
DEFINITION H.sapiens mRNA for nicotinic acetylcholine receptor alpha7 subunit
precursor.
X08420
ACCESSION Y2808623
NID Y08420.1 GI:2808623
VERSION nAChR gene; nicotinic acetylcholine receptor alpha 7 subunit.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1509)
AUTHORS Groot Kormelink, P.J. and Luyten, W.H.
TITLE Cloning and sequence of full-length cDNAs encoding the human
neuronal nicotinic acetylcholine receptor (nAChR) subunits betas3
and beta4 and expression of seven nAChR subunits in the human
neuroblastoma cell line SH-SY5Y and/or IMR-32
FEB8 Lett. 400 (3), 309-314 (1997)
MEDLINE 97162233
REFERENCE 2 (bases 1 to 1509)
AUTHORS Groot Kormelink, P.J.
TITLE Direct Submision
JOURNAL Submitted (27-SEP-1996) P.J. Groot Kormelink, Janssen Research
Foundation, Exp. Mol. Biol. Dept., Turnhoutseweg 30, B-2340 Beerse,
Belgium
REMARK Revised by author 22-JAN-1998
COMMENT On Jan 25, 1998 this sequence version replaced gi:1702915.
FEATURES
source location/Qualifiers
1..1509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="neuroblastoma"
/cell_line="IMR-32"
/cell_line="SH-SY5Y"
1..66
/gene="nAChRA7"
1..1509
/gene="nAChRA7"
/codon_start=1
/product="nicotinic acetylcholine receptor alpha7 subunit
precursor"
/protein_id="CAA69697.1"
/db_xref="PID:e1246208"
/db_xref="PID:g2808624"
/db_xref="GI:2808624"
/translation="MRCSPGGVWLALASLLHVSLSGSEFORLKYELKYNPNLEERV
ANDQPLTVFSLIQIMDVDEKNQVLTINIMLQMSMTDHYLQMNVSSEYPGKTVRF
PDGQIMKPDILLYSADERPDATFHTNVLVNSGHCQYLPGRFSSCYIDVWPEPD
VQHKLFKFSWSTGWSLQMDQADISGYPNGEMDVLGIPGRSRSEYECCKEYRP
DVTFTVMRRRTLYGNTLIPCVLISALALVFLPADSGEKISGIVTLSTVFM
LVLAELMPATSDSVPLIAQYFASMTIMVGLSVVTVIVLQYHHHDPGGMKPKWTWYI
LNMCAWFLRMKRGEDKVRPAQHKORCSLASVMSAVAPPPASNGNLVIGFRGL
DVHCVPTPDSDGVYCGRMACSPTHDEHLHGQPPEDPDIAKILIEVRYIANRFRCO
DESEAVCSSEMKFACVYDRLCLMAFSVFTICTIGILMSAPNFEAVNSKQFA"
1..1509
/gene="nAChRA7"
mat_peptide 67..1506
/gene="nAChRA7"
BASE COUNT 297 a 451 c 429 g 332 t
ORIGIN
Query Match 15.7%; Score 108; DB 10; Length 1509;

Best Local Similarity 95.7%; Pred. No. 1,6e-16;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 574 CTGCGATATGCGATGAGAGCGCTTTGAGCCACATTCACAGTGTGGTAATTC 633
Db 342 CTATACAGTCTGATGAGCGCTTTGACGCCACATTCACACTACAGCTGTGTAATTC 401
OY 634 TTCTGGGCAATGGCAGTACCTGCGCCAGCATATTCAGAGTCTCGCTACATCG 689
Db 402 TTCTGGGCAATGGCAGTACCTGCGCCAGCATATTCAGAGTCTCGCTACATCG 457

RESULT 15
LOCUS HSU62436 1876 bp mRNA PRI 11-JAN-1997
DEFINITION Human nicotinic acetylcholine receptor alpha7 subunit precursor,
mRNA, complete cds.
U62436
ACCESSION U62436
NID 91458119
VERSION U62436.1 GI:1458119
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1876)
AUTHORS Elliott, K.J., Ellis, S.B., Berckhan, K.J., Urrutia, A.,
Chavez-Noriega, L.E., Johnson, E.C., Velicelebi, G. and Harpold, M.M.
TITLE Comparative structure of human neuronal alpha 2-alpha 7 and beta
2-beta 4 nicotinic acetylcholine receptor subunits and functional
expression of the alpha 2, alpha 3, alpha 4, alpha 7, beta 2, and
beta 4 subunits
J. Mol. Neurosci. 7 (3), 217-228 (1996)
MEDLINE 97062879
REFERENCE 2 (bases 1 to 1876)
AUTHORS Elliott, K.J.
TITLE Direct Submision
JOURNAL Submitted (28-JUN-1996) Kathryn J. Elliott, SIBRA Neurosciences,
Inc., 505 Coast Blvd. So., La Jolla, CA 92037, USA
FEATURES
source location/Qualifiers
1..1876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="SIBRA lambda gtl1 library #3 (M. Williams)"
/cell_line="IMR32"
/cell_type="neuroblastoma"
1..72
/gene="nAChRA7"
73..138
/product="nicotinic acetylcholine receptor alpha7 subunit
precursor"
/protein_id="AAB40114.1"
/db_xref="PID:g1458120"
/db_xref="GI:1458120"
/translation="MRCSPGGVWLALASLLHVSLSGSEFORLKYELKYNPNLEERV
ANDQPLTVFSLIQIMDVDEKNQVLTINIMLQMSMTDHYLQMNVSSEYPGKTVRF
PDGQIMKPDILLYSADERPDATFHTNVLVNSGHCQYLPGRFSSCYIDVWPEPD
VQHKLFKFSWSTGWSLQMDQADISGYPNGEMDVLGIPGRSRSEYECCKEYRP
DVTFTVMRRRTLYGNTLIPCVLISALALVFLPADSGEKISGIVTLSTVFM
LVLAELMPATSDSVPLIAQYFASMTIMVGLSVVTVIVLQYHHHDPGGMKPKWTWYI
LNMCAWFLRMKRGEDKVRPAQHKORCSLASVMSAVAPPPASNGNLVIGFRGL
DVHCVPTPDSDGVYCGRMACSPTHDEHLHGQPPEDPDIAKILIEVRYIANRFRCO
DESEAVCSSEMKFACVYDRLCLMAFSVFTICTIGILMSAPNFEAVNSKQFA"
139..1578
/product="nicotinic acetylcholine receptor alpha7 subunit"
mat_peptide 139..1578
BASE COUNT 369 a 553 c 531 g 423 t
ORIGIN
Query Match 15.7%; Score 108; DB 10; Length 1876;

Best Local Similarity 95.7%; Pred. No. 1.6e-16;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	574	CTCGGATATGCTGATGAGCGCTTTGAGCGCCACATTCACACTAACGTGTGTGAATTC	633
Db	414	CTATACAGTGTGATGAGCGCTTTGAGCGCCACATTCACACTAACGTGTGTGAATTC	473
QY	634	TTCTGGGCAATGGCAGTACCTGCTCCAGGCAATTCAGAGTCTCTGCTACATCG	689
Db	474	TTCTGGGCAATGGCAGTACCTGCTCCAGGCAATTCAGAGTCTCTGCTACATCG	529

Search completed: September 17, 1999, 22:05:19
Job time: 16502 sec

1	108	15.7	1876	1	T48239
2	108	15.7	1876	1	V12197
3	108	15.7	1876	1	V44687
4	74.4	10.8	2769	1	T59196
5	61.6	8.9	2101	1	T59197
6	46.8	6.8	5561	1	V68408
7	46.8	6.8	5561	1	V68409
8	46	6.7	114955	1	X58491
9	44.6	6.5	2460	1	Q06086
10	42.8	6.2	2383	1	T95785
11	41.2	6.0	2106	1	X28103
12	40.2	5.8	2992	1	T67283
13	40.2	5.5	2990	1	V66832
14	38.2	5.5	4488	1	T35520
15	38	5.5	1578	1	T35520
16	37	5.4	1128	1	Q61817
17	36.6	5.3	3223	1	T74074
18	36.6	5.3	3223	1	Q62917
19	36.4	5.3	1915	1	T48241
20	36	5.2	114955	1	X53491
21	36	5.2	4020	1	T91361
22	35.4	5.1	1120	1	V60339
23	35.4	5.1	1120	1	V60339
24	35.4	5.1	11213	1	Q64888
25	35.2	5.1	2529	1	V62176
26	35	5.1	4897	1	Q03259
27	35	5.1	1915	1	V12198
28	34.8	5.1	2058	1	X06987
29	34.6	5.0	12001	1	G76213
30	34.4	5.0	2363	1	V12196
31	34.4	5.0	201	1	N70191
32	34.4	5.0	1536	1	Q44491
33	34.4	5.0	3496	1	T48237
34	34.4	5.0	1925	1	T48234
35	34.4	5.0	2082	1	T59527
36	34	4.9	2082	1	T59528
37	33.8	4.9	774	1	T44439
38	33.8	4.9	1018	1	T02336
39	33.8	4.9	1018	1	T87014
40	33.6	4.9	43280	1	T80413
41	33.6	4.9	2721	1	Q10212
42	33.6	4.9	2721	1	Q10543
43	33.6	4.9	3000	1	Q14319
			3000	1	Q34566
					prn gene. DNA enco

08-MAR-1993; US-028031.
(SALK) SALK INST BIOTECHNOLOGY IND ASSOC.

PI Elliott KJ, Ellis SB, Harpold MM;
DR WPI: 94-303024/37.
DR P-PSDB: W44153.
PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -
PT also transformed cells useful for screening cpds. which modulate
PT activity of the receptor
PS Claim 8: Page 78-79; 99pp; English.
CC The present sequence encodes a human neuronal nicotinic acetylcholine
CC receptor (NACHR) subunit. The cells expressing the alpha and/or beta
CC NACHR subunits may be used in a method of screening compounds to
CC identify any which modulate the activity of human neuronal NACHR.
CC Subunit specific antibodies may be used to monitor the distribution
CC and expression density of various subunits in normal vs diseased brain
CC tissues. Testing of single receptor subunits or specific receptor
CC subunit combinations with a variety of potential agonists or antagonists
CC provides information with respect to the function and activity of the
CC individual subunits and should lead to the identification and design of
CC compounds that are capable of very specific interaction with one or
CC more receptor subtypes. The resulting drugs should exhibit fewer
CC unwanted side effects than drugs identified e.g. screening with cells
CC that express a variety of subtypes.
CC Sequence 1876 BP; 369 A; 553 C; 530 G; 423 T;
Query Match 15.7%; Score 108; DB 1; Length 1876;
Best Local Similarity 95.7%; Pred. No. 4.8e-21;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCACATTCACACTAGCGTTGGTGAATTC 633
DB 414 CTATACAGTGTGATGAGCGCTTTGAGCCACATTCACACTAGCGTTGGTGAATTC 473
OY 634 TTCTGGGCAATTCGCCAGTACCTGCTCCAGGCATATTCAAGAGTTCTCTACATCG 689
DB 474 TTCTGGGCAATTCGCCAGTACCTGCTCCAGGCATATTCAAGAGTTCTCTACATCG 529
RESULT 3
ID V44687 standard; cDNA; 1590 BP.
AC V44687;
DE 09-OCR-1998 (first entry)
DE V274T variant human alpha7 nACHR coding sequence.
KW Alpha7 nACHR; alpha7 nicotinic acetylcholine receptor subunit; cancer;
KW neurodegeneration; enzyme dysfunction; affective disorder; therapy;
KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;
KW schizophrenia; ss.
OS Homo sapiens.
Key Location/Qualifiers
CDS 9..1517
F1 /*tag- a
F1 WO9828331-A2.
F1 02-JUL-1998.
F1 22-DEC-1997; U23405.
F1 20-DEC-1996; US-771737.
F1 (ABSO) ABBOTT LAB.
F1 Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,
F1 Roch J, Sullivan JP, Touma E;
F1 WPI: 98-377593/32.
F1 P-PSDB: W69216.
CC Nucleic acid encoding variant of human alpha7 nicotinic
CC acetylcholine receptor sub-unit - used to identify modulators of
CC the receptor, potentially useful for treating neuro-degeneration,
CC cancer, affective disorders etc.
PS Claim 14: Fig 2; 44pp; English.
CC This sequence encodes the V247T variant of human alpha7 nicotinic
CC acetylcholine receptor (NACHR) subunit of the invention. Cells containing
CC the DNA are used to express the protein and to identify modulators of
CC alpha7 nACHR activity or cytoprotective agents, e.g. antisense
CC compounds or antagonists that are potentially useful for treating
CC neurodegeneration, enzyme dysfunction, affective disorders and immune
CC dysfunction, such as cancer, post-herpetic neuralgia, diabetic
CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,

CC psychosis and schizophrenia. Probes based on the DNA are used to detect
CC the DNA in usual hybridization or amplification tests, while monoclonal
CC antibodies are used to detect the protein for diagnosis (in vitro or by
CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nACHR,
CC the protein has about 100-fold greater sensitivity to cholinergic
CC receptor agonists (nicotine or acetylcholine) and response to these
CC agonists decays more slowly, but the wild-type inward rectification is
CC retained.
CC Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;
Query Match 15.7%; Score 108; DB 1; Length 1590;
Best Local Similarity 95.7%; Pred. No. 4.5e-21;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCACATTCACACTAGCGTTGGTGAATTC 633
DB 350 CTATACAGTGTGATGAGCGCTTTGAGCCACATTCACACTAGCGTTGGTGAATTC 409
OY 634 TTCTGGGCAATTCGCCAGTACCTGCTCCAGGCATATTCAAGAGTTCTCTACATCG 689
DB 410 TTCTGGGCAATTCGCCAGTACCTGCTCCAGGCATATTCAAGAGTTCTCTACATCG 465
RESULT 4
ID T59196 standard; cDNA; 2769 BP.
AC T59196;
DE 17-JUN-1997 (first entry)
DE Neuronal alpha-bungarotoxin binding protein alpha subunit cDNA.
DE Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
KW ligand binding; ion channel; ss.
OS Gallus sp.
FH Key Location/Qualifiers
FT cgs 71..1513
FT signal_peptide 71..136
FT /*tag- b
FT mat_peptide 137..1510
FT /*tag- c
PN US5599709-A.
PD 04-FEB-1997.
PF 28-SEP-1989; 413947.
PR 28-SEP-1989; US-413947.
PI (SALK) SALK INST BIOLOGICAL STUDIES.
PI Lindstrom JM, Schoepfer RD;
DR WPI: 97-118297/11.
DR P-PSDB: W12368.
PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
PT to screen cholinergic agents and other drugs which may affect ligand
PT binding; ion channel or other activities of the protein.
PS Claim 1; Fig 2A-B; 18pp; English.
PS 2 cDNA clones (T59196 and T59197), respectively code for the alpha1
CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
CC an 18-day embryo chick brain cDNA library using a probe (see also
CC T59198) based on the N-terminal amino acid sequence of chicken
CC brain ABBP. The probe isolated partial clone pch29-1, which
CC encoded the N-terminal portion of alpha1. A subclone, pch29-3
CC (ATCC 40641), was used to rescreen the library, yielding clone
CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
CC A probe based on the C-terminal region of pch31-1 was used obtain
CC clone pch34-1 (ATCC 40639), encoding the C-terminal portion of
CC alpha1. The cDNA clones can be used as probes to identify further
CC ABBP subunits, and in the recombinant prodn. of ABBP.
CC Sequence 2769 BP; 790 A; 604 C; 571 G; 804 T;
Query Match 10.8%; Score 74.4; DB 1; Length 2769;
Best Local Similarity 77.6%; Pred. No. 1.5e-11;
Matches 90; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
OY 574 CTGCATATTGCTGATGAGCGCTTTGAGCCACATTCACACTAGCGTTGGTGAATTC 633

Db 412 CTGTAACAGCTGCTGATGAAGATTGATGCTACATTTCACACTAATGTTTAGTCAATTC 471
 QY 634 TTCTGGGCAATGCGGAGTACCTGCTCCAGGCAATTCACAGTTCCTCGTACATCG 689
 Db 472 TTGCGGACACTGCGCAATATCTGCGACAGGCAATTTAAAGCTCATGCTACATAG 527

RESULT 5

TS9197
 ID T59197 standard; cDNA; 2101 BP.
 AC T59197;
 DT 17-JUN-1997 (first entry)
 DE Neuronal alpha-bungarotoxin binding protein alpha2 subunit cDNA.
 KW Neuronal alpha-bungarotoxin binding protein alpha 2; cholinergic;
 KM ligand binding; ion channel; ss.
 OS Gallus sp.
 Key cds Location/Qualifiers
 cds 56..1501
 FT signal_peptide 56..145 /*tag= a
 FT mat_peptide 146..1498 /*tag= b
 FT /*tag= c
 PN US559709-A.
 PD 04-FEB-1997.
 PF 28-SEP-1989; 413947.
 PR 28-SEP-1989; US-413947.
 PA (SALK) SALK INST BIOLOGICAL STUDIES.
 PI Lindstrom JM Schoepfer RD;
 DR WPI: 97-118297/11.
 DR P-PSDB; W12369.
 PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
 PT binding, ion channel or other activities of the protein.
 PS Claim 1; Fig 3A-B; 18pp; English.
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
 CC an 18-day embryo chick brain cDNA library using a probe (see also
 CC T59198) based on the N-terminal amino acid sequence of chicken
 CC brain ABBP. The probe isolated partial clone pCh29-1, which
 CC encoded the N-terminal portion of alpha1. A subclone, pCh29-3
 CC (ATCC 40641), was used to rescreen the library, yielding clone
 CC pCh31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
 CC A probe based on the C-terminal region of pCh31-1 was used obtain
 CC clone pCh34-1 (ATCC 40639), encoding the C-terminal portion of
 CC alpha1. The cDNA clones can be used as probes to identify further
 CC ABBP subunits, and in the recombinant prodn. of ABBP.
 SQ Sequence 2101 BP; 582 A; 406 C; 454 G; 659 T;

Query Match
 Best Local Similarity 8.9%; Score 61.6; DB 1; Length 2101;
 Matches 82; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 574 CTGCGATATGCTGATGAGGCGTTTGACGCGCATTCACACTAAGTGTGGTGAATTC 633
 Db 421 CTATACAGTGGCGATGAAGATTGATGCAATTCACACAAAGTGTGATTA 480
 QY 634 TTCTGGGCAATGCGGAGTACCTGCTCCAGGCAATTCAGAGTTCCTGCTACATCG 689
 Db 481 CTTCTGATCTCTGCAATATATTCCTCCAGGCAATTTGAAGACACATGTTCAATTG 536

RESULT 6

V68408/c
 ID V68408 standard; cDNA to mRNA; 5561 BP.
 AC V68408;
 DT 05-MAY-1999 (first entry)
 DE Human BAZ1-beta cDNA #1.
 KW Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ;
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;

KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 346..4929
 FT /*tag= a
 FT /*product= "BAZ1-beta"

PN W09847920-A1.

PD 29-OCT-1998.
 PF 17-APR-1998; J01783.
 PR 24-OCT-1997; JP-310027.
 PR 18-APR-1997; JP-116570.
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PI Jones MH;
 DR WPI: 98-583603/49.
 DR P-PSDB; W81172.
 PT Transcriptional regulator gene family containing bromodomain - may
 PT be expressed in testis tissue and is useful for treatment of cancer
 PT and other proliferative disorders
 PS Claim 2; Page 125-137; 187pp; Japanese.
 CC This sequence encodes the human BAZ1-beta protein, a member of a
 CC family of transcriptional regulator genes containing a bromodomain (BAZ,
 CC in testis tissue and also in certain tumour lines. Transgenic cells may
 CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
 CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
 CC and other proliferative disorders, and in screening of compounds for
 CC their binding ability to the expression products (e.g. for use as drugs
 CC by modulation of transcriptional regulation).
 SQ Sequence 5561 BP; 1685 A; 1261 C; 1471 G; 1164 T;

Query Match
 Best Local Similarity 6.8%; Score 46.8; DB 1; Length 5561;
 Matches 111; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 10 CAAAGCGGTAGCGGGGCGGAGTGTGCTGCTGCTTTAAAGTGGGCTGACGGAGCC 69
 Db 264 CAAAGCGCGGGGGGTGGGTGGGAGAGGAGGAGGAGGCGGCGGAATCTCGGC 205
 QY 70 GCGGCTCTCTGCGTGTGAGTGTGATTAAGGAGGAGCGCCGAGCGGCAATAGCT 129
 Db 204 TCCCTCAAGCCCGCGCGCGCGGCGGACAGTATGAGGCGGAGGCGGCGGCAAGCT 145
 QY 130 CCGGCAAGTCTCGGTGGGCGGCTTGCATTTTCACCGCGCTCCACAGAGGTCAAGCG 189
 Db 144 CCGGCGACACCGCGCGGCTCCACAGACCGCCCGGACCTCCGCTTGGGTCCCGC 85
 QY 190 GCGGGGAGAGGTGAGCGCGGAGAGCTGCGCGGGGG 227
 Db 84 GCGGCGAGTCAAGACTCTCTCTGAGACAGCGGAGG 47

RESULT 7

V68409/c
 ID V68409 standard; cDNA to mRNA; 5573 BP.
 AC V68409;
 DT 05-MAY-1999 (first entry)
 DE Human BAZ1-beta cDNA #2.
 KW Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ;
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
 KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 346..4941
 FT /*tag= a
 FT /*product= "BAZ1-beta"
 PN W09847920-A1.
 PD 29-OCT-1998.
 PF 17-APR-1998; J01783.
 PR 24-OCT-1997; JP-310027.
 PR 18-APR-1997; JP-116570.
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PI Jones MH;

DR WP1: 98-583603/49.
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PS Claim 2; Page 145-157; 187pp; Japanese.

CC This sequence encodes the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc Finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).

SQ Sequence 5573 BP; 1668 A; 1267 C; 1472 G; 1166 T;

Query Match 6.8%; Score 46.8; DB 1; Length 5573;
Best Local Similarity 50.9%; Pred. No. 0.001;
ches 111; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

DY 10 CCAGGCGGTACGGGGGGCAATGCTCCTTGTGCCCTTTAAACTGGGGCTTGACGGAGACC 69
DB | | | | | | | | | | | | | | | | | | | |
DY 264 CGAGCGCCCCGGGGGGTGGGGGAAGAAGGGGGGTGAGAGSGGGGCAGACTCCGGC 205
DB | | | | | | | | | | | | | | | | | | | |

DY 70 GCAGCTTCCTGTCGTGAGTCGATTATAAAGGACAGCACGCCGCGACGCACATAGCT 129
DB | | | | | | | | | | | | | | | | | | | |

DY 204 TCCTCACCAGCGCGGGGGCCGCGGACAGTATGAGAGCGACGCCACGGCGAGAGCT 145
DB | | | | | | | | | | | | | | | | | | | |

DY 130 CCCGCAAAGCTCTCGTGGTCCCCTTGGCATTTTTCAGCGCGGCTCCAGAGGGTCAAGGC 189
DB | | | | | | | | | | | | | | | | | | | |

DY 144 CCGCGGACACCGCGCGGCTCCAGACAGCCCCCGCGAGCTCCGCTTGCGGTCCCGGC 85
DB | | | | | | | | | | | | | | | | | | | |

OY 190 GGCGGGGAGAGGTGAGAGCCCGAGAGCTCGCGGGGG 227
DB ||| |||| || | | | | | | | | | |

DB 84 GCGCGGAGTCAAGCACTCTCTCCAGAGACAGCGAGG 47
DB ||| |||| || | | | | | | | | | |

RESULT 8
X53491 AC Human adenosine A1 receptor antisense oligonucleotide fragment.
X53491 DT 05-JUL-1999 (first entry)
KW Antisense oligonucleotide; multiple target; antisense treatment;
impaird respiration; inflammation; lung disease;
pulmonary vasoconstriction; inflammation; allergic rhinitis;
acute asthma; allergy; asthma; impeded respiration;
respiratory distress syndrome; pain; cystic fibrosis;
pulmonary hypertension; pulmonary vasoconstriction; emphysema;
chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
colon cancer; breast cancer; lung cancer; pancreatic cancer; melanoma;
hepatocellular carcinoma; kidney cancer; metastasis;
prostate cancer; ss.
Synthetic.
OS WO913886-A1.
PM 25-MAR-1999.
PD 17-SEP-1998; U19419.
PF 09-JUN-1998; US-093972.
PR 17-SEP-1997; US-059160.
PA (UYEC-) UNIV EAST CAROLINA.
PI Nycex JW;
WP1: 99-229400/19.
DR New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction
PS Disclosure; Page 37; 120pp; English.
CC The specification describes antisense oligonucleotides (X52869-X53271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene
CC initiation codons, genomic flanking regions, intron-exon borders, the
CC 5'-end, the 3'-end and the junction between coding and non-coding
CC regions and all segments of RNAs encoding proteins associated with one

	CC	or more diseases, conditions or mixtures. The antisense oligonucleotides
	CC	may be derived from sequences X5272-74. These multiple target
	CC	oligonucleotides (specifically X5280-271) can be used for the antisense
	CC	treatment of diseases and conditions. Typical diseases and conditions
	CC	are those associated with impaired respiration and inflammation,
	CC	including lung diseases, pulmonary vasoconstriction, inflammation,
	CC	allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
	CC	respiratory distress syndrome, pain, cystic fibrosis, pulmonary
	CC	hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
	CC	pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
	CC	cervical cancer, breast cancer, lung cancer, pancreatic
	CC	cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
	CC	metastases, as well as all types of cancers which may metastasize or have
	CC	metastasized to the lungs, including breast and prostate cancer.
	SQ	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;
Query Match	6.7%; Score 46; DB 1; Length 114955;	
Best Local Similarity	39.4%; Pred. No. 0.0054;	
Matches 114; Conservative 22; Mismatches 151; Indels 2; Gaps 1.		
QY	9	CCGACGGAGTGTGGCGGAGTCGTGTCTGTGCCTTTAACTGCGCTTGACGGAGAC 68
Db	104584	CGCCGCCGCCGCCSNNNDNBGGCCBGGCGCGCCGCCGCCGCCSNNNDNNGCC.105043
		: :
QY	69	C GCACCCTCTGTCTGCTGTGTGAGTCTGATTATTAAGAAGAGACAGCCCGCAGGCC--GCCACATA 126
Db	105044	BGGCGGGCGCGCGCGCGCGCGCGSNNNDNNGCCBGGCGCGCGCGCGCGCGCGCGSNNN 105103
		: :
QY	127	GCTCCGCCCAAGTCTCTGTGTGTGCTCCCTTGTCCATTTCACAGCGCGCTCCACAGAGGTAC 186
Db	105104	DNNCCBGGCGCGCGCGCGCGCGCGSNNNDNBGGCBGGCGCGCGCGCGCGCGCGS 105163
		: :
QY	187	GCGCGGGGAGAGTAGTGGAGCGCGGAGAGTCTGCGCGGGGGGCCCGCTGTGTGGCGCGG 246
Db	105164	NNNDNBGGGGCGCGCGCGCGCGCGSNNNDNBGGGGCGCGCGCGCGCGCGSNN 105223
		: :
QY	247	CCAAGACAGCGGCTCGGAGACTGTCTTTTCGCGCCCTCCCGCGG 295
Db	105224	NNDNNGCGCGCGCGCGCGCGCGCGSNNNDNBGGCGCGCGCGCGCGCGCG 105272
		: :
RESULT	9	
ID	Q06086	
AC	O06086 standard; cDNA; 2460 BP.	
DT	24-JAN-1991 (first entry)	
DE	Plasmid pZPC13 encoding neuronal nicotinic acetylcholine receptor	
DE	beta 4 subunit.	
KW	Rat; NACHR; ss.	
OS	Rattus rattus.	
FH	Key	
FT	cds	Location/Qualifiers
FT	cds	61..1548
FT		/tag- a
FT	signal_peptide	61..120
FT		/tag- b
FT	mat_peptide	121..1545
FT		/tag- c
FT	misc_feature	/product=NACHR beta 4 subunit
FT		416..417
FT		/tag- d
FT	misc_feature	/label=splice site
FT		1409..1410
FT	misc_feature	/tag- e
FT		/label=splice site
FT		1569..1612
FT		/tag- f
FT	repeat_unit	/number=1
FT		1625..1668
FT		/tag- g
FT	repeat_unit	/number=2
FT		1683..1726
FT		/tag- h
FT	repeat_unit	


```

FT      /number-3
FT      2436..2441
FT      /*tag-1
FT      conflict
FT      replace(268..270,gac)
FT      /*tag-1
FT      /note="conflict between genomic and cDNA sequences"
PN      MO9010648-A.
PD      20-SEP-1990.
PF      14-MAR-1990; U01403.
PR      14-MAR-1989; US-321384.
PA      12-MAR-1990; US-492555.
PI      (SALK ) SALK INST FOR BIOL STUD.
PI      Heilmann SF; Denertis ES; Duvoisin RM; Patrick JW;
DR      WPI; 90-304987/40.
DR      P-PSDB; R07143.
PT      New neuronal nicotinic acetylcholine receptor - compans. contg.
PT      beta 4 sub-unit and DNA sequences encoding them.
PT      Claim 11: Fig 3; 47pp; English.
CC      The sequence encodes a novel neuronal nicotinic acetylcholine
CC      receptor subunit, beta 4. Plasmid pPC13 (ATCC 67893) was
CC      isolated from a cDNA library in lambda zap II prep. from rat
CC      mRNA. The sequence of the cDNA in the region encoded by the 5th
CC      exon is identical to that determined for the genomic clone with
CC      the exception of a subst. at posn. 720 of a t for a c (this does
CC      not alter the AA sequence); this may be due to a polymorphism
CC      between the rat strains used for the genomic library and the cDNA
CC      library, or could have resulted from a reverse transcriptase error
CC      during prep. of the cDNA. A sequence, the core of which is 44 bp
CC      long is repeated three times at the beginning of the 3' UT region.
CC      The function of this remains unknown. The protein subunit can
CC      combine with the known subunits, alpha-2,-3, and -4 and beta-2
CC      to form previously unknown functional receptors. The new subunit
CC      is expressed in the central and peripheral nervous systems and in
CC      PC12 cells.
SQ      Sequence 2460 BP; 519 A; 767 C; 587 G; 587 T;

Query Match
Best Local Similarity 6.5%; Score 44.6; DB 1; Length 2460;
Matches 71; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY      575 TGCATATTCGTATGACGCGTTGACCCACATTCACATCAAGTGTGTTGTAATCT 634
DB      409 TACACACATGCGGATGACCATATGAGTGTCTCTACACACATGATGTCCTCC 468
C        635 TCTGGGCTTCCGCTACCTCCCTCCAGGCAATATCAAGAGTCTCTACATCG 689
L        469 AACGGCAGCATTCAGTGGCTGCCCTCTCTATCTACAAAGAGTGCCTGCAAGATTG 523

RESULT 10
T95785 standard; cDNA to mRNA; 2383 BP.
AC      T95785;
DT      22-MAY-1998 (first entry)
DE      Maize starch synthase type I cDNA
KW      Maize; starch synthase type I; starch; ds.
OS      Zea mays.
FS      Key
FH      Location/Qualifiers
FT      2..1951
FT      /*tag-a
FT      /product= starch_synthase_type_I
PD      DE19619918-A1.
PD      20-NOV-1997.
PF      17-MAY-1996; 019918.
PR      17-MAY-1996; DE-019918.
PA      (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PI      Froberg C; Kossmann J;
PI      WPI; 98-000821/01.
DR      P-PSDB; W38218.
DR      DNA encoding maize starch synthase type I protein - for producing
PT      transgenic plants
PT      Claim 1; Pages 16-20; 23pp; German.

```

```

CC      The present sequence encodes maize starch synthase type I, useful
CC      in the production of starch. Starch can be used in various
CC      conventional starch applications, e.g. starch hydrolysate products,
CC      foods, papermaking, adhesives, textiles, building materials, soil
CC      stabilisation, agrochemicals, pharmaceuticals, cosmetics, coal
CC      briquettes, ore and coal slurries, foundry casting, rubber, leather
CC      and synthetic polymers. The enzyme produces a starch stated to have
CC      different physicochemical properties, especially viscosity and
CC      gelling properties, from wild type starch.
SQ      Sequence 2383 BP; 562 A; 546 C; 633 G; 622 T;

Query Match
Best Local Similarity 6.2%; Score 42.8; DB 1; Length 2383;
Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY      170 GCTCCACGAGGCTACGCGCGGAGAGGTGAGCCGAGAGCTGGCGGGGGCC 229
DB      121 GCTCCAGCGCGTGTCTCGCGCGCGGTGCTCCGAGAGTGAAGAGGCGCCCGCC 180
QY      230 CGGCTGTGGCGCGCGCCATGACAGCGGCTCTCTTTTCGCGCCCTCC 289
DB      181 GCGCCGATGCCACCGCGCTCTGCGCCCGCGCTCGTCCGCGCTCTCTGCGCGCC 240
QY      290 CCGCGAGGTGAGGGAAGATGTCATGTC 319
DB      241 GGCAGGCGCCAGCGGTGAGCGCGCATGAC 270

RESULT 11
X28103/c
ID      X28103 standard; DNA; 2106 BP.
AC      X28103;
DT      11-JUN-1999 (first entry)
DE      Frecac11 gene.
KW      Frecac11; transcription factor; adipose tissue expressed gene; obesity;
KW      lipid metabolism; adipocyte differentiation; obesity-related condition;
KW      non-insulin dependent diabetes mellitus; cardiovascular disease;
KW      catabolic condition; anorexia; bulimia; therapy; ss.
OS      Homo sapiens.
PN      W09854216-A1.
PD      03-DEC-1998.
PF      26-MAY-1998; SE0989.
PR      26-MAY-1997; SE-001963.
PA      (PMAA ) PHARMACIA & UPJOHN AB.
PI      Carlsson P; Enerbaeck S;
PI      WPI; 99-070211/06.
DR      P-PSDB; Y01097.
PT      New transcription factor designated FRECAC11 which regulates
PT      adipose tissue expressed genes involved in lipid metabolism and
PT      adipocyte differentiation, used to, e.g. inhibit gene expression
PS      Claim 1; Fig 1; 23pp; English.
CC      This sequence represents the Frecac11 gene of the invention. The encoded
CC      protein has transcriptional regulatory function directed against adipose
CC      tissue expressed genes. The protein is involved in lipid metabolism
CC      and/or adipocyte differentiation. FRECAC11 is selectively expressed in
CC      adipose tissue, and is used for transcriptional regulation of adipocyte
CC      obesity-related conditions to increase or decrease the activity of
CC      adipocyte expressed genes, e.g. FRECAC11 can be used as a drug that blocks
CC      the cis elements of adipose expressed genes to inhibit gene expression.
CC      The FRECAC11 protein can also be used to down regulate screening for
CC      substances that affect the activity of FRECAC11, such as inhibitors,
CC      antagonists, or agonists. Conditions which can be treated include
CC      obesity, non-insulin dependent diabetes mellitus, cardiovascular
CC      diseases, catabolic conditions, anorexia, bulimia.
SQ      Sequence 2106 BP; 419 A; 757 C; 602 G; 328 T;

Query Match
Best Local Similarity 6.0%; Score 41.2; DB 1; Length 2106;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```


CC constructs useful for gene activation by homologous recombination.
 CC Novel genes are generated in which exogenous and endogenous exons
 CC are operatively linked. Expression of IPO encoded by the endogenous
 CC gene is activated, providing a means for the gene therapy of
 CC thrombocytopenia.
 SQ Sequence 4488 BP; 875 A; 1440 C; 1439 G; 722 T;

Query Match
 Best Local Similarity 47.3%; Score 38.2; DB 1; Length 4488;
 Matches 115; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 60 GACGGAGCCGCGCTCTCTGCTGAGTGTATTAAGGAGCAAGCCCGAGCCG 119
 DB 2401 GCGGAGAGGCGCGGCGCGGCGGCGGAGGAGGCGGCGGCGGCGGCGG 1111
 QY 120 CCAATAGCTCCCGCCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460
 2461 GGGCAATGCTCCCGAGAGGCGCTGCTGCGGCGCGGAGAGTCCGCGCGGCG 179
 QY 180 GGGTCAAGGCGGCGGAGAGTGTGAGCCGAGAGCTGCGCGGCGCGCGGCGG 2520
 DB 2521 CCGGAGAGAGGCGGCGGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 239
 QY 240 GCCGCGCCATGACAGCGGCTGCGGACTGCTCTTTCGCGCGCGCGCGGAGGT 299
 DB 2581 CTACTGCCCCCGGCGCGGCTCGGCGCTTTATGCCCCGCGGCGGCGGCGG 2640
 QY 300 GAG 302
 DB 2641 GGG 2643

RESULT 15

Q61817
 ID Q61817 standard; DNA; 1578 BP.
 AC Q61817;
 DT 11-NOV-1994 (first entry)
 DE Open reading frame of pseudorabies virus.
 KW Pseudorabies virus; antigen; polypeptide; vaccine; pig; swine;
 OS disease; mutant; mutation; ds.
 PS Pseudorabies virus.
 FH Key
 FT cds
 FT 1.1578
 FT Location/Qualifiers
 FT /tag- a
 FT /product- Antigenic or functional polypeptide of
 FT pseudorabies virus.

FN W09408000-A.
 PD 14-APR-1994.
 PF 06-OCT-1993; E02738.
 PA 06-OCT-1992; EP-203079.
 PA (ALKU) AK20 NY.
 PI Mettenleiter TC.
 DR WPI; 94-135567/16.
 DR P-PSDB; R51272.
 PT New pseudo:rabies virus mutants - having a mutation in an open
 PT reading frame, used partic. in vaccines for immunising pigs.
 PS Claim 1; Page 23-24; 33pp; English.
 CC A pseudorabies virus with a mutation in this open reading frame can
 CC be used in the production of vaccines for combatting infectious
 CC diseases in pigs. The pseudorabies virus mutant produced fails to
 CC produce the antigenic or functional polypeptide described in R51272
 CC and expresses a polypeptide with altered antigenic or functional
 CC properties or one that is non-functional. The mutation is either a
 CC nucleic acid substitution, deletion, insertion or inversion or a
 CC combination of these. The preferred substitutions are a combination
 CC of substitutions selected from those where the nucleic acid is
 CC modified so that the codon encoding His at position 37 is replaced
 CC by one encoding Arg, the codon encoding Glu at position 35 is
 CC changed to one encoding Asp, and the codon encoding Val at position
 CC 375 is changed to a codon encoding Ala. A mutation in the control
 CC elements of the ORF localised in an intergenic sequence may also
 CC result in a failure of expression of the polypeptide.

SQ Sequence 1578 BP; 186 A; 568 C; 607 G; 217 T;

Query Match
 Best Local Similarity 50.5%; Score 38; DB 1; Length 1578;
 Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 53 GCGGCTGAGGAGGAGCCGCGCTCTCTGCTGAGTGTATTAAGGAGAGCCCG 112
 DB 1019 GCGGCGGCGGCGGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGG 1078
 QY 113 CAGCGCCGACATAGCTCCCGCAATCTCTGCTGCTGCTGCTGCTGCTGCT 172
 DB 1079 GCGGCGGAGGCGGCTCTCGGCGCGGAGAGCGCTGCGCCGCAACCGCGTCT 1138
 QY 173 CCGACGAGGATCACGCGGCGGAGAGAGTGGAGCCGCGAGAGCTGCGCGGCGG 232
 DB 1139 GCGAGCGAGCGCTCTGCGGCGGAGAGTGTATGCGGATGCTCTGTAACCGCG 1198
 QY 233 CC 234
 DB 1199 CC 1200

Search completed: September 18, 1999, 00:35:12
 Job time: 19072 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 21:28:33 ; Search time 2825.05 Seconds

(without alignments)
447.565 Million cell updates/sec

Title: US-08-956-518a-103

Perfect score: 641
Sequence: 1 CAGGCCCGCCACATAGCTGCC.....CAGAGTTCCTGCTACATCG 641

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	151.8	23.7	560	26	W52861	W52861 zc03e02.r1
2	130.8	20.4	409	25	W03952	W03952 za62c08.r1
3	106	16.5	170	25	N83152	N83152 K4329F Huma
4	92.2	14.4	510	54	HSW011948	A1047098 Homo sapi
5	76.4	11.9	487	39	AA861176	AA861176 ak36d09.s
6	68.6	10.7	335	32	AA333244	AA333244 EST37317
7	49.4	7.7	607	44	A1292581	A1292581 GH1518.5
8	46	7.2	501	49	A1655990	A1655990 tt42c03.x
9	43.6	6.8	274	34	AA460018	AA460018 zx49h09.s
10	43.6	6.8	436	39	AA844642	AA844642 at57a12.s
11	43.6	6.8	396	45	A1367541	A1367541 qv93c12.x
12	40.8	6.4	494	37	AA697326	AA697326 H02253.5
13	39.4	6.1	439	47	A1492967	A1492967 q247c02.x
14	39.2	6.1	354	26	W86937	W86937 zh60e03.r1
15	39	6.1	557	43	A1192683	A1192683 g62d07.x
16	38.6	6.0	369	29	AA156428	AA156428 z151c02.s
17	37.8	5.9	790	48	A1595069	A1595069 vb77d03.y
18	37	5.7	473	33	AA407186	AA407186 EST02181
19	36.6	5.7	319	33	AA408177	AA408177 EST02759
20	36.6	5.7	433	40	AA923335	AA923335 O144e04.s
21	36.4	5.7	400	50	A1681298	A1681298 tx45f10.x
22	36.4	5.7	399	51	AU067986	AU067986 AU067986
23	36	5.6	435	42	A1085561	A1085561 cy68a04.x
24	36	5.6	344	45	A1611153	A1611153 qy05c04.x
25	36	5.6	373	47	A1462417	A1462417 ub62e06.x
26	36	5.6	373	47	A1467007	A1467007 ub78b04.x
27	35.6	5.5	415	24	H81299	H81299 yu74f03.s1
28	35.4	5.5	317	46	A1443300	A1443300 sa46b11.y
29	35.2	5.5	638	26	W27222	W27222 26b9 Human
30	35.2	5.5	654	45	A1346522	A1346522 gp51d09.x
31	35	5.5	232	20	D22064	D22064 RICC10276A
32	35	5.5	232	20	D22108	D22108 RICC10350A
33	35	5.5	216	20	D24865	D24865 RICR2674A.R
34	35	5.5	482	39	AA894793	AA894793 c155e11.s
35	35	5.5	482	40	A1660583	A1660583 we68d07.x
36	34.6	5.4	368	40	AA953883	AA953883 on76d09.s
37	34.4	5.4	328	33	AA407004	AA407004 EST02243
38	34.4	5.4	416	33	AA407437	AA407437 EST00734
39	34.4	5.4	417	43	A1198318	A1198318 q161g07.x
40	34.4	5.4	461	43	A1199677	A1199677 q160a10.x
41	34.4	5.4	351	45	A1364148	A1364148 qy76f02.x
42	34.4	5.4	495	48	A1564797	A1564797 tn37d02.x
43	34.4	5.4	372	50	AU064505	AU064505 AU064505
44	34.2	5.3	328	24	H86162	H86162 ys94b03.r1
45	34.2	5.3	298	43	A1227041	A1227041 u131h12.y

ALIGNMENTS

RESULT 1
LOCUS W52861
DEFINITION zc03e02.r1 Soares.Parathyroid_tumor_NHPR Homo sapiens cDNA clone
IMAGE:321242 5 similar to gb:X70297 NEURONAL ACETYLCHOLINE
RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN);, mRNA sequence.
ACCESSION W52861
NID g1350351

VERSION W52861.1 GI:1350351
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 560)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT On Jan 24, 1995 this sequence version replaced gi:634363.
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1753 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 408.
 Location/Qualifiers
 1..560
 /organism="Homo sapiens"
 /db_xref="GDB:1258900"
 /db_xref="taxon:9606"
 /clone="IMAGE:321242"
 /clone_lib="Soares parathyroid tumor_NbHRA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: parathyroid gland; Vector: p773D
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 15'-
 TGTTCACATCTGAGTGGAGCGCGCCACCAATTTTCTTTTCTTTTCTTTT
 T-3', double-stranded cDNA was size selected, ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified p773
 vector (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
 adenomas was kindly provided by Dr. Stephen Marx, National
 Institute of Diabetes and Digestive and Kidney Diseases,
 NIH."

BASE COUNT 128 a 137 c 137 g 152 t 6 others
 ORIGIN

Query Match 23.7%; Score 151.8; DB 26; Length 560;
 Best Local Similarity 96.4%; Pred. No. 5.1e-29;
 Matches 186; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 449 AATTGATGCAAAATATGTCATCTACACGATTTTCATTCGTAATCCAGCAT 508
 Db 1 ATTTTGTGCAAAAATATGTCATCTACACGATTTTCATTCGTAATCCAGCAT 60
 QY 509 TTGTGATACCTCAAACTGCGATATGTCATGAGCGCTTGAGCCCATTCACACT 568
 Db 61 TTGTGATACCTCAAACTGCGATATGTCATGAGCGCTTGAGCCCATTCACACT 119
 QY 569 AACGCTGTGGAATTTCTTGTGGGCAATGGCAGTACCTGCTCCAGGACATTCAGAGT 628
 Db 120 AACGCTGTGGAATTTCTTGTGGGCAATGGCAGTACCTGCTCCAGGACATTCAGAGT 177
 QY 629 TCCTGCTACATCG 641

Db 178 TCCTGCTACATCG 190
 RESULT 2
 LOCUS W03952
 DEFINITION W03952 409 bp mRNA EST 19-APR-1996
 2662C08.r1 Soares fetal liver spleen INFIS Homo sapiens cDNA clone
 IMAGE:297134 5' similar to gb:X70297 NEURONAL ACETYLCHOLINE
 RECEPTOR PROTEIN, ALPHA-7 CHAIN (HUMAN); mRNA sequence.
 ACCESSION W03952
 MID G1275817
 VERSION W03952.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 409)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: mob.REGA+ET
 High quality sequence stop: 263.
 Location/Qualifiers
 1..409
 /organism="Homo sapiens"
 /db_xref="GDB:1242055"
 /db_xref="taxon:9606"
 /clone="IMAGE:297134"
 /clone_lib="Soares fetal liver spleen INFIS"
 /dev_stage="20 week-post conception fetus"
 /sex="male"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 15' AACCTGAGAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTTCTTTT
 T-3', double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 95 a 80 c 75 g 145 t 14 others
 ORIGIN

Query Match 20.4%; Score 130.8; DB 25; Length 409;
 Best Local Similarity 92.4%; Pred. No. 1.1e-23;
 Matches 146; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 482 TTTCGATTCATTCGTAATCA-6CATTGTGATGTCAGTCAAACTGATATGCTGA 540
 Db 133 TTTCGATTCATTCGTAATTCACGACATTTTNGATGCTGAATTCGATATGCTGA 212
 QY 541 TGAGGCTTTGAGCGCATTCACACACTAACGTTGTTGTAATTTCTTGTGGGCAATGCCA 600
 Db 213 TGAGGCTTTGAGCGCATTCACACACTAACGTTGTTGTAATTTCTTGTGGGCAATGCCA 272
 QY 601 GTACCTGCTCCAGGACATTCACAGATTCCTGCTACA 638

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 487)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced gi:937098.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.linn.gov/dbp/IMAGE/IMAGE.html

Insert Length: 758 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 458.
Location/Qualifiers
1. 487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1408049"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'.
TGTTACCAATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT733 vector. Library
went through one round of normalization to Cots, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 95 c 102 g 159 t
ORIGIN

Very Match 11.9%; Score 76.4; DB 39; Length 487;
Best Local Similarity 98.7%; Pred. No. 8.7e-10;
Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 457 GCAAAATATTTGCATCTCCAGCATTTTCAGTTCGAATTCGAATCCAGCATTTGTGAT 516
|||
Db 487 GCAAAATATTTGCATCTCCAGCATTTTCAGTTCGAATTCGAATCCAGCATTTGTGAT 428
|||
Y 517 AGCTGCAAACTGCATAT 534
|||
Db 427 AGCTGCAAACTGTGATAT 410
|||

RESULT 6
AA333244
LOCUS
DEFINITION AA333244 335 bp mRNA EST 21-APR-1997
AA333244
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE
AUTHORS
Eutheria: Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 335)
Adams,M.D., Kellavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Linn,N.H., Kirkness,E.F., Weinstock,K.G., Gockayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georgagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedlund,E., Hinkle,P.S.J.,
Kellely,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spilgus,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Fertile,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melsner,P.S., Olsen,H.,
Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On May 18, 1995 this sequence version replaced gi:811121.
COMMENT
JOURNAL
MEDLINE
Other_ESTS: THC185696
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018698056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/ngi/ngi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES
source
1..335
/organism="Homo sapiens"
/db_xref="rnc:(inhost):134920"
/db_xref="taxon:9606"
/clone_lib="Embryo, 8 week 1"
/dev_stage="embryo, 8 wks"
/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-";
Site1: EcoRI; Site2: XhoI"
BASE COUNT
84 a 94 c 88 g 69 t
ORIGIN
Query Match 10.7%; Score 68.6; DB 32; Length 335;
Best Local Similarity 84.6%; Pred. No. 7.5e-08;
Matches 77; Conservative 0; Mismatches 14; Indels 0; Gaps 0.
QY 175 TCCGCGCGAGTGAGGGAGAGATGTCATGTCAGAGGTCAGAGCCAAAGCAACTACT 234
|||||
Db 1 TCCGCGCGAGATGAGGGAGAGATGTCGCTGTCAGAGCGCCAGAGTGAAGTTCCT 60
QY 235 GGCGTCATCTTCCAGGAGAACCCAGAGCCA 265
|||||
Db 61 GGCGTCATCTTCCAGGAGAACCCAGAGCGA 91
RESULT 7
LOCUS AI292581 607 bp mRNA EST 01-DEC-1998
DEFINITION GH1518, 5prime GH Drosophila melanogaster head P072 Drosophila
melanogaster cDNA clone GH1518 5prime, mRNA sequence.
ACCESSION AI292581
NID 93941988
VERSION AI292581.1 GI:3941988
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 607)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brockstein, P., Lewis, S. and Rubin, G.M.
BDGP/HHMI Drosophila EST Project
Unpublished (1997)
On Jun 18, 1996 this sequence version replaced gi:1366631.

FEATURES
source

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
hit genomic sequence AC005890
Plate: 155 row: B column: 6
High quality sequence stop: 465.
Location/Qualifiers
1..607
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH15518"
/clone_1lb="GH Drosophila melanogaster head POT2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="organ: head; Vector: POT2; Site:1; EcorI; Site:2; XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 157 a 151 c 162 g 137 t
ORIGIN

Query Match 7.7%; Score 49.4; DB 44; Length 607;
Best Local Similarity 64.3%; Pred. No. 0.0073;
Matches 74; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 527 TGCGATATTGCTGATGAGCGTTGACGCACATTCACACTAGCTGTGGAATTC 586
DB 398 TACACAGCGCGGATGACGATTCGATGACGATCACCACCGCTTGCGCAACAT 457
QY 587 TGTGGCATGTGCCAGTACCTGCTCCAGCATATTCAGAGATTCCGTACATCG 641
DB 458 GCGCGAGTGTCTGTACGTGCCCGCTGATCTCAAGACGACATCAAGATGG 512

LOCUS 8
DEFINITION A1655990 501 bp mRNA EST 04-MAY-1999
ACCESSION U42C03.x1 NCI_CGAP_G6 Homo sapiens cDNA clone IMAGE:2243428 3',
NID A1655990
VERSION 94739966
KEYWORDS A1655990.1 GI:4739969
SOURCE EST.
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
On Jun 5, 1998 this sequence version replaced gi:13188057.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/btrp/image/image.html

Seq primer: -40bp from glbco
High quality sequence stop: 458.
Location/Qualifiers
1..501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2243428"
/clone_1lb="NCI_CGAP_G6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_G64 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 107 c 115 g 144 t
ORIGIN

Query Match 7.2%; Score 46; DB 49; Length 501;
Best Local Similarity 83.9%; Pred. No. 0.051;
Matches 52; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 427 ATGAACACACCCATCGTTAAATTGATGCAAAATTTTCATCTACACATTTCA 486
DB 499 ATAAAGAAACACCCACCATTAATTGATGCAAAATTTTCATCTACACATTTCA 440
QY 487 GT 488
DB 439 GT 438

RESULT 9
LOCUS AA460018/c
DEFINITION AA460018 274 bp mRNA EST 09-JUN-1997
ACCESSION ZK49H03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:795617
NID AA460018
VERSION 92184902
KEYWORDS AA460018.1 GI:2184902
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Hillier, L., Allen, M., Bowles, L., Dubbuge, T., Gelsel, G., Jost, S., Kueba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Scheinberg, R., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
JOURNAL WasNU-Merck EST Project 1997
COMMENT Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1393691.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevatscn.wustl.edu
This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -41ml3 fwd. ET from Amersham.
 Location/Qualifiers

FEATURES
 SOURCE

1. 274
 /organism="Homo sapiens"
 /db_xref="GDB:6038561"
 /db_xref="taxon:9606"
 /clone="IMAGE:795617"
 /clone_1lb="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5].
 TGTTCACCAATGCTAATCCAGCATTTGTGATAGCTGCAACTGCGATAT 534
 1 ||||||||||||||||||| ||||||| |||||
 Db 258 CTGTCCAAATGCTAATCCAGCATTTGTGATAGCTGCAACTGCGATAT 209
 1 ||||||||||||||||||| ||||||| |||||
 Query Match 6.8%; Score 43.6; DB 34; Length 274;
 Best Local Similarity 92.0%; Pred. No. 0.17; 4; Indels 0; Gaps 0;
 Matches 46; Conservative 0; Mismatches 0; Gaps 0;
 BASE COUNT 78 a 53 c 67 g 76 t
 ORIGIN

Query Match 6.8%; Score 43.6; DB 34; Length 274;
 Best Local Similarity 92.0%; Pred. No. 0.17; 4; Indels 0; Gaps 0;
 Matches 46; Conservative 0; Mismatches 0; Gaps 0;
 BASE COUNT 78 a 53 c 67 g 76 t
 ORIGIN

RESULT 10
 AA844642 436 bp mRNA EST 31-DEC-1998
 LOCUS a157a12.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394398
 DEFINITION 3', mRNA sequence.
 ACCESSION AA844642
 NID 52931093
 VERSION AA844642.1 GI:2931093
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 436)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900948.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html

FEATURES
 SOURCE
 Insert Length: 497 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 415.
 Location/Qualifiers
 1. 436
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/map="1.1"
 /clone="IMAGE:1394398"
 /clone_1lb="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5].
 TGTTCACCAATGCTAATCCAGCATTTGTGATAGCTGCAACTGCGATAT 534
 1 ||||||||||||||||||| ||||||| |||||
 Db 269 CTGTCCAAATGCTAATCCAGCATTTGTGATAGCTGCAACTGCGATAT 220
 1 ||||||||||||||||||| ||||||| |||||
 Query Match 6.8%; Score 43.6; DB 39; Length 436;
 Best Local Similarity 92.0%; Pred. No. 0.2; 4; Indels 0; Gaps 0;
 Matches 46; Conservative 0; Mismatches 0; Gaps 0;
 BASE COUNT 115 a 96 c 103 g 122 t
 ORIGIN

Query Match 6.8%; Score 43.6; DB 39; Length 436;
 Best Local Similarity 92.0%; Pred. No. 0.2; 4; Indels 0; Gaps 0;
 Matches 46; Conservative 0; Mismatches 0; Gaps 0;
 BASE COUNT 115 a 96 c 103 g 122 t
 ORIGIN

RESULT 11
 A1367541 396 bp mRNA EST 15-FEB-1999
 LOCUS qv93c12.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:1989142 3',
 DEFINITION mRNA sequence.
 ACCESSION A1367541
 NID 94137286
 VERSION A1367541.1 GI:4137286
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html

FEATURES
 SOURCE
 Insert Length: 1828 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 370.
 Location/Qualifiers
 1. 396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1989142"
 /clone_1lb="NCI_CGAP_Ut2"
 /tissue_type="moderately-differentiated endometrial
 adenocarcinoma, 3 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;

Site 2: NotI: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 98 a 86 c 97 g 115 t

Query Match 6.8%; Score 43.6; DB 45; Length 396;
Best Local Similarity 92.0%; Pred. No. 0.19;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 485 CAGTCCATTGCTATCCAGCATTTGTGATGCGCAACGCGATAT 534
1 |||||
Db 280 CTGTCATTATGCTATCCAGCATTTGCGATAGCTCCAACTGGGATAT 231

RESULT 12
AA697326 494 bp mRNA EST 28-NOV-1998
LOCUS HL02253.5prime HL Drosophila melanogaster head Bluescript
DEFINITION M20316: D.melanogaster acetylcholine receptor-related protein mRNA,
complete cds, mRNA sequence.

ACCESSION AA697326
NID 92700255
VERSION AA697326.1 GI:2700255
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 494)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brostein, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
On Nov 20, 1997 this sequence version replaced gi:1172316.

COMMENT
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSH, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 22 row: E column: 5
High quality sequence stop: 416.

FEATURES
source Location/Qualifiers
1..494
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="HL02253"
/clone_1lb="HL Drosophila melanogaster head Bluescript"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/note="Organ: head-brain & sensory organ; Vector:
Bluescript SK; Site 1: EcoRI; Site 2: XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK(+/-)."

BASE COUNT 124 a 114 c 127 g 129 t

Query Match 6.4%; Score 40.8; DB 37; Length 494;
Best Local Similarity 59.5%; Pred. No. 1.1;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 526 CTGCGATATGCTGATGACGCTTTGACGCGACATTCACACTACGCTGTGATGCAATTC 585
1 |||||
Db 244 CTGATATATGCGGATGCGACACTACGAGGTGGCGCTCAAGTCCAAAGCTGCTATTAATCC 303

QY 586 TTCTGGCATTCGCCAGTACCTGCTCCAGCATATTCAGAGCTTCTGTACATCG 641
1 |||||
Db 304 CACGGGAGAGGTCTGTGGTCTCTCCGCGCATTTACAGAGACTCTGCACATCG 359

RESULT 13
A1492967 439 bp mRNA EST 18-MAR-1999
LOCUS q247c02.x1 NCI-CGAP_Ki411 Homo sapiens cDNA clone IMAGE:2030018 3'
DEFINITION similar to SW:CDNC_HUMAN P49918 CYCLIN-DEPENDENT KINASE INHIBITOR
1C; contains TAR1.t2 TAR1 repetitive element;; mRNA sequence.

ACCESSION A1492967
NID 94393970
VERSION A1492967.1 GI:4393970
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 439)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3136536.

COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLM at:
www.bio.livl.gov/dbtrp/image/image.html

FEATURES
source Location/Qualifiers
1..439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q22.1-3"
/clone="IMAGE:2030018"
/clone_1lb="NCI-CGAP_Ki411"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonids 1323376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 60 a 145 c 187 g 47 t

Query Match 6.1%; Score 39.4; DB 47; Length 439;
Best Local Similarity 53.3%; Pred. No. 2.3;
Matches 105; Conservative 0; Mismatches 91; Indels 1; Gaps 1;

QY 26 GTCCTGATGCCCCCTTCATTTCCAGCCGCGCTCCACGAGGCTCAGCGGCGGGGA 85
1 |||||
Db 46 GGCAGCGGCGCCCGTGGTTCGGCGCTCTGCTTGAAGCCCGCGGTCCGGGG 105
1 |||||
QY 86 GAGGTGAGACCCCGAGAGCTCGCGGGGCGCCGCTGTGTGGCGCGCATGACAGCG 145
1 |||||

D _b	106	CCGAGGGCGGGGAGCGGGAGGCCGGGGACCGGGGCGGGT-CTGTGGGGCCGGGGCCGC	164
O _y	146	GTCGCGGACTGCGTCTTTTTCGCGGCCCTTCCGCGCGAGAGTGAGGGAGAACATGTTCATG	205
D _b	165	ACTGGAACCGCGGCGCGGAGCCGAGACCGGAGCGCGGSCCGGGGACCAGACCGAG	224
O _y	206	TCAAGGTTCAAGGCCAA	222
D _b	225	ACCGCGACCGGAGCCGA	241

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 354)	Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Holtman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Marids, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Roifling, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, D., Trevasaks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, and Marra, M.	Generation and analysis of 280,000 human expressed sequence tags	Genome Res. 6 (9), 807-828 (1996)	
57044478				On Apr 14, 1993 this sequence version replaced g1:315330.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAG Consortium (info@image.lnl.gov) for further information.
Insert Length: 584 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 217.

FEATURES
SOURCE

```

/organism="Homo sapiens"
/db_xref="GDB:1324946"
/db_xref="taxon:9606"
/clone="IMAGE:416475"
/clone_id="Soares-fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; site_1: Pac I; site_2: Eco RI
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACCTGAGAGATTTATTTAATGAAGCTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."

```

Query Match	6.1%	Score 39.2;	DB 26;	Length 354;
Best Local Similarity	75.3%	Pred. No. 2.5		
Matches 61;	Conservative	0;	Indels 19;	Gaps 1;
QY 485	CAGTTCACATTCGTAATCAGCATTTGGATGCTGCAGAACTGGATATTGCTATGAG	544		
DB 86	CTGTTCATATTGCTATATCAGATTGCGGANA-NTCCAAACTGGGATATTGTCTCAGG	144		
QY 545	CGCTTGACGCCACATTCAC	565		
DB 145	CGCTGTGCTTCACCTTCAC	165		

RESULT	15
LOCUS	Al192683
DEFINITION	Al192683 557 bp mRNA EST 29-OCT-1998
	g662807.x1 Soares fetal_lung Homo sapiens CDNA clone
	IMAGE:1743565.3, similar to SW:CDKN HUMAN P4918 CYCLIN-DEPENDENT
	KINASE INHIBITOR 1C; contains MER22.b3 TAA1 repetitive element ;
	mRNA sequence.
ACCESSION	Al192683
NID	93743892
VERSION	Al192683.1 GI:3743892
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 557)
AUTHORS
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 14, 1998 this sequence version replaced gi:1797633.

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNCI; contact the
IMAG Consortium (info@imgc.llnl.gov) for further information.
Insert length: 862 Std Error: 0.00
Seq primer: -40bp from Gbpco
High quality sequence stop: 118.
Location/Qualifiers
1..557
source

```

/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Lung; Vector: pRT3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
5'-GTGTCCCAATCTGAAGTGGCAGCGCCGCAATTTTTTTTTTTT-3'}
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT3 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Benito
Soares and M.Fatima Bomaldi. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NDH19W."

```

[illegible]

Db 68 GGGCGGCTGCCCTGCTGCTGCGGCCCTGCTCGGCGCTCTCTTGAGGCGCCGGCTCCGGG 127
QY 86 GAAGTGGAGCCCGGAGACTCGGCGGGGGCCCGCTGCTGAGCCGCGGCATGACAGCG 145
Db 128 CCGGGGCCCGGGGCGGGGCGCGGGCCCGGGGCGG-3GCTGGGCGCGGGCGCGG 186
QY 146 GCTCGGAGCTGCTCTTTCCGCGGCCCTCCGCGCGGAGTGAGGGGAAGATGTCCATG 205
Db 187 ACTGGAGCCGCGGCGGAGCCGAGCCGAGCCGCGGGCGCGGGCCGAGGACCGCG 246
QY 206 TCAGGTTCAAGCC 220
Db 247 ACCGCGACCGAGCC 261

Search completed: September 17, 1999, 21:28:34
Job time: 14298 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 00:35:12 ; Search time 425.19 Seconds

(without alignments)
377.180 Million cell updates/sec

Title: US-08-956-518A-103

Perfect score: 641
Sequence: 1 CAGGCCGCCACATAGCTCC.....CAGAGTTCCTGCTACATCG 641

Scoring table: IDENTITY_NUC

arched: 311585 segs, 125096042 residues

_cabase : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	16.8	1876	1 T48239	Neuronal nicotinic
2	108	16.8	1876	1 V12197	Human neuronal nic
3	108	16.8	1590	1 V44687	V274T variant huma
4	74.4	11.6	2769	1 T59196	Neuronal alpha-bun
5	61.6	9.6	2101	1 T59197	Neuronal alpha-bun
6	44.6	7.0	2460	1 Q06086	Plasmod parp13 enc
7	43	6.7	11955	1 X53491	Human adenosine A1
8	42.8	6.7	2383	1 T95785	Maize starch synth
9	41.2	6.4	2106	1 X28103	Freacil gene. New
10	40.2	6.3	2992	1 T67285	Soluble starch syn
11	40.2	6.3	2990	1 V66832	Zea mays soluble s
12	36.6	5.7	1915	1 T48241	Neuronal nicotinic
13	36.4	5.7	5561	1 V68408	Human BAZ1-beta CD
14	35.4	5.5	1942	1 Q84888	Human BAZ1-beta CD
15	35.4	5.5	1915	1 V12198	DNA encoding cytop
16	35	5.5	2538	1 X06987	Human neuronal nic
17	35	5.5	2538	1 V12196	Mouse neuregulin r
18	34.6	5.4	2363	1 V12196	Human neuronal nic
19	34.4	5.4	201	1 N70195	Streptomyces prote
20	34.4	5.4	1536	1 Q44491	Therius aquaticus
21	34.4	5.4	4488	1 T35520	Human thrombopoiet
22	34.4	5.4	3496	1 T48237	Neuronal nicotinic
23	34.4	5.4	1925	1 T48234	Neuronal nicotinic
24	34.4	5.4	2082	1 T59527	Alpha4 subunit of
25	34.4	5.4	2082	1 T59528	Alpha4 subunit of
26	34.4	5.4	1120	1 V60339	CDNA sequence of f
27	34	5.3	3223	1 Q62917	hnmnp u protein co
28	34	5.3	774	1 T44499	Human cyclin-depo
29	33.8	5.3	1018	1 T02326	Human cardiac hype
30	33.8	5.3	1018	1 T87014	Human cardiac hype
31	33.8	5.3	43280	1 T80413	Tyrosine synthase
32	33.6	5.2	2721	1 Q10212	BamHI J-I fragment
33	33.6	5.2	2721	1 Q10543	BamHI J-I fragment
34	33.6	5.2	3000	1 Q14319	Pectactin antigen
35	33.6	5.2	3000	1 Q34566	prn gene. DNA enco
36	33.6	5.2	8438	1 Q73500	DNA encoding Pseud
37	33.6	5.2	1597	1 T12908	Human insulin rece
38	33.6	5.2	4739	1 T78203	Physiologically ac
39	33.6	5.2	8169	1 V26609	Actinomadura hibis
40	33.4	5.2	201	1 N70194	Signal portion of
41	33.4	5.2	2064	1 Q52638	Streptomyces fradi
42	33.4	5.2	2821	1 T89389	Human cystatin B g
43	33.4	5.2	2821	1 T89394	Human mutant EPM1

ALIGNMENTS

C	44	33.4	5.2	2821	1	T89395	Mutant EPM1 gene A
.45	33.4	5.2	1320	1	V44138	Chimeric rHPDGF-B	
ALIGNMENTS							
RESULT	1						
ID	T48239						
AC	T48239, standard; DNA; 1876 BP.						
DT	09-APR-1997 (first entry)						
DE	Neuronal nicotinic acetylcholine receptor alpha-7 subunit DNA.						
KW	Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;						
OS	Homo sapiens.						
FH	Key						
FT	cds						
FT	cds						
FN	W09641876-A1.						
PD	27-DEC-1996.						
PF	07-JUN-1996; U09775.						
PR	07-JUN-1995; US-484722.						
PA	(SIBI-) SIBIA NEUROSCIENCES INC.						
PI	Elliot KJ, Harpold KM;						
DR	WPI; 97-065463/06.						
DR	P-PSDB; W09025.						
PT	Nucleic acids encoding neuronal nicotinic acetylcholine receptor sub-units -						
PS	used in screening to determine the effect of drugs on the receptor						
PS	Disclosure; Page 71-73; 108pp; English.						
CC	A DNA sequence (T48239) codes for the alpha-7 subunit (W09025) of						
CC	the human neuronal nicotinic acetylcholine receptor (nAChR). Host						
CC	cells, esp. mammalian cells or amphibian oocytes, carrying alpha-7						
CC	nucleic acids, opt. in combination with other alpha and/or beta						
CC	subunit nucleic acids (see also T48232-38, T48240-41), express						
CC	recombinant nAChR subunits useful for identifying cpds. that						
CC	modulate the activity of human nAChRs.						
CC	Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;						
SO	Sequence 1876 BP; 369 A; 553 C; 531 G; 423 T;						
Query Match 16.8%; Score 108; DB 1; Length 1876;							
Best Local Similarity 95.7%; Pred. No. 9.1e-23;							
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;							
QY	526	CTGCATATTTGCTGATGAGCGTTGAGCCACATTCACACTACAGTGGTGAATTC	585				
DB	414	CTATACAGCTGCTGATGAGCGTTGAGCCACATTCACACTACAGTGGTGAATTC	473				
QY	586	TTCTGGGATTCGCCAGTACCTGCTCCAGGATATTCAGAGTTCCTGCTACATCG	641				
DB	474	TTCTGGGATTCGCCAGTACCTGCTCCAGGATATTCAGAGTTCCTGCTACATCG	529				
RESULT 2							
ID	V12197						
AC	V12197 standard; cDNA; 1876 BP.						
DT	14-MAY-1998 (first entry)						
DE	Human neuronal nicotinic acetylcholine receptor alpha-7 subunit cDNA.						
KW	Human; neuronal nicotinic acetylcholine receptor; alpha-7 subunit;						
OS	Brain tissue; screening; NACHR; antibody; ds.						
OS	Homo sapiens.						
FH	Key						
FT	cds						
FT	cds						
FN	W09420617-A2.						
PD	15-SEP-1994.						
PF	08-MAR-1994; U02447.						
PR	08-MAR-1993; US-028031.						
PA	(SALK) SALK INST BIOTECHNOLOGY IND ASSOC.						
PA	(SIBI-) SIBIA NEUROSCIENCES INC.						

PI Elliott KJ, Ellis SB, Harpold KM;
 DR WPI: 94-303024/37.
 DR P-PSDB: W44153.
 PT Human neuronal nicotinic acetylcholine receptor subunits and DNA -
 CC also transformed cells useful for screening cpds. which modulate
 PT activity of the receptor
 PS Claim 8; Page 78-79; 99pp; English.
 CC The present sequence encodes a human neuronal nicotinic acetylcholine
 CC receptor (NACHr) subunit. The cells expressing the alpha and/or beta
 CC NACHr subunits may be used in a method of screening compounds to
 CC identify any which modulate the activity of human neuronal NACHr.
 CC Subunit specific antibodies may be used to monitor the distribution
 CC and expression density of various subunits in normal vs diseased brain
 CC tissues. Testing of single receptor subunits or specific receptor
 CC subunit combinations with a variety of potential agonists or antagonists
 CC provides information with respect to the function and activity of the
 CC individual subunits and should lead to the identification and design of
 CC compounds that are capable of very specific interaction with one or
 CC more receptor subtypes. The resulting drugs should exhibit fewer
 CC unwanted side effects than drugs identified e.g. screening with cells
 CC that express a variety of subtypes
 Sequence 1876 BP; 369 A; 533 C; 530 G; 423 T;
 Query Match 16.8%; Score 108; DB 1; Length 1876;
 Best Local Similarity 95.7%; Pred. No. 9.1e-23;
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 526 CTGCGATATGCTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585
 DB 414 CTATACAGTGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 473
 QY 586 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGATTCCTGTACATCG 641
 DB 474 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGATTCCTGTACATCG 529
 RESULT 3
 V44687
 ID V44687 standard; cDNA; 1590 BP.
 AC V44687;
 DT 09-OCT-1998 (first entry)
 DE V274T variant human alpha7 NACHr coding sequence.
 KW Alpha7 NACHr; alpha7 nicotinic acetylcholine receptor subunit; cancer;
 KW neurodegeneration; enzyme dysfunction; affective disorder; therapy;
 KW immune dysfunction; diabetic neuropathy; Alzheimer's disease;
 KW schizophrenia; ss.
 OS Homo sapiens.
 Key Location/Qualifiers
 CDS 9..1517
 /tag= a
 PI W09828331-A2.
 PN 02-JUL-1998.
 PR 22-DEC-1997; U23405.
 PR 20-DEC-1996; US-771737.
 PA (ABBO) ABBOT LAB.
 PI Briggs CA, Gopalakrishnan M, McKenna DG, Monteggia LM,
 PI Roch J, Sullivan JP, Touna E;
 DR WPI: 98-377593/32.
 P-PSDB: W69216.
 PT Nucleic acid encoding variant of human alpha7 nicotinic
 PT acetylcholine receptor sub-unit - used to identify modulators of
 CC the receptor, potentially useful for treating neuro-degeneration,
 CC cancer, affective disorders etc.
 PS Claim 14; Fig 2; 44pp; English.
 CC This sequence encodes the V247T variant of human alpha7 nicotinic
 CC acetylcholine receptor (NACHr) subunit of the invention. Cells containing
 CC the DNA are used to express the protein and to identify modulators of
 CC alpha7 NACHr activity or cytoprotective agents, e.g. antisense
 CC compounds or antagonists that are potentially useful for treating
 CC neurodegeneration, enzyme dysfunction, affective disorders and immune
 CC dysfunction, such as cancer, post-herpetic neuralgia, diabetic
 CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, kuru,

CC psychosis and schizophrenia. Probes based on the DNA are used to detect
 CC the DNA in usual hybridisation or amplification tests, while monoclonal
 CC antibodies are used to detect the protein for diagnosis (in vitro or by
 CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 NACHr,
 CC the protein has about 100-fold greater sensitivity to cholinergic
 CC receptor agonists (nicotine or acetylcholine) and response to these
 CC agonists decays more slowly, but the wild-type inward rectification is
 CC retained.
 Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;
 Query Match 16.8%; Score 108; DB 1; Length 1590;
 Best Local Similarity 95.7%; Pred. No. 8.4e-23;
 Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 526 CTGCGATATGCTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585
 DB 350 CTATACAGTGTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 409
 QY 586 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGATTCCTGTACATCG 641
 DB 410 TTCTGGGCAATTCGACGACTGCTGCTCCAGGCAATTCACAGATTCCTGTACATCG 465
 RESULT 4
 T59196
 ID T59196 standard; cDNA; 2769 BP.
 AC T59196;
 DT 17-JUN-1997 (first entry)
 DE Neuronal alpha-bungarotoxin binding protein alpha1 subunit cDNA.
 KW Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
 KW ligand binding; ion channel; ss.
 OS Gallus sp.
 FH Gallus sp.
 FT cds
 FT key Location/Qualifiers
 FT 71..1513
 FT signal_peptide 71..136
 FT mat_peptide 137..1510
 FT /tag= b
 FT /tag= c
 PN US5599709-A.
 PD 04-FEB-1997.
 PR 28-SEP-1989; 413947.
 PR 28-SEP-1989; US-413947.
 PA (SALK) SALK INST BIOLOGICAL STUDIES.
 PI Lindstrom JM, Schoepfer RD;
 DR WPI: 97-118297/11.
 P-PSDB: W12368.
 DT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
 PT to screen cholinergic agents and other drugs which may affect ligand
 PT binding, ion channel or other activities of the protein.
 PS Claim 1; Fig 2A-B; 18pp; English.
 CC 2 cDNA clones (T59196 and T59197) respectively code for the alpha1
 CC alpha-bungarotoxin binding protein (ABBP) of chick neuronal
 CC an 18-day embryo chick brain cDNA library using a probe (see also
 CC T59198) based on the N-terminal amino acid sequence of chicken
 CC brain ABBP. The probe isolated partial clone pCh29-1, which
 CC encoded the N-terminal portion of alpha1. A subclone, pCh29-3
 CC (ATCC 40641), was used to rescreen the library, yielding clone
 CC pCh31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
 CC A probe based on the C-terminal region of pCh31-1 was used obtain
 CC clone pCh34-1 (ATCC 40639), encoding the C-terminal portion of
 CC alpha1. The cDNA clones can be used as probes to identify further
 CC ABBP subunits, and in the recombinant prodn. of ABBP.
 Sequence 2769 BP; 790 A; 604 C; 571 G; 804 T;
 Query Match 11.6%; Score 74.4; DB 1; Length 2769;
 Best Local Similarity 77.6%; Pred. No. 1.4e-12;
 Matches 90; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 526 CTGCGATATGCTGATGAGCGCTTTGACGCCACATTCACACTACAGTGTGGTGAATTC 585

DB 412 CTATACACAGTCTGATGGAAGATTTATGCTTACATTCACACATTAAGTTTATGTAATTC 471
 OY 586 TTCTGGGACATTCGACGATGACGCTTCACGACATATTCAGAGTCTGCTAGATCG 641
 DB 472 TTCTGGGACATTCGACGATGACGCTTCACGACATATTCAGAGTCTGCTAGATCG 527

RESULT 5

TS9197
 ID TS9197 standard; cDNA; 2101 BP.
 AC TS9197;

DE 17-UN-1997 (first entry)
 DE Neuronal alpha-bungarotoxin binding protein alpha2 subunit cDNA.
 KM Neuronal alpha-bungarotoxin binding protein alpha 2; cholinergic;
 OS ligand binding; ion channel; ss.

Key cds Location/Qualifiers
 56..1501

FT signal_peptide 56..145
 FT mat_peptide 146..1498
 FT /*tag= a
 FT /*tag= b
 FT /*tag= c

PD 04-FEB-1997.
 PD 28-SEP-1989; 413947.
 PR 28-SEP-1989; US-413947.
 PA (SALK) SALK INST BIOLOGICAL STUDIES.
 PI Lindstrom JM. Schoepfer RD.
 DR WPI: 97-118297/11.
 DR P-PSDB: W12369.

PT New isolated neuronal alpha-bungarotoxin-binding protein DNA - used
 PT to screen cholinergic agents and other drugs which may affect ligand
 PT binding, ion channel or other activities of the protein.
 PS Claim 1: Fig 3A-B; 18pp; English.

CC 2 cDNA clones (T59196 and T59197), respectively code for the alpha1
 CC subunit (W12368) and alpha2 subunit (W12369) of chick neuronal
 CC alpha-bungarotoxin binding protein (ABBP). They were isolated from
 CC an 18-day embryo chick brain cDNA library using a probe (see also
 CC T59198) based on the N-terminal amino acid sequence of chicken
 CC brain ABBP. The probe isolated partial clone pch29-1, which
 CC encoded the N-terminal portion of alpha1. A subclone, pch29-3
 CC (ATCC 40641), was used to rescreen the library, yielding clone
 CC pch31-1 (ATCC 40640), which encoded the entire alpha2 sequence.
 CC A probe based on the C-terminal region of pch31-1 was used to obtain
 CC clone pch4-1 (ATCC 40639), encoding the C-terminal portion of
 CC alpha1. The cDNA clones can be used as probes to identify further
 CC ABBP subunits and in the recombinant prodn. of ABBP.
 CC Sequence 2101 BP; 582 A; 406 C; 454 G; 659 T;

Query Match 9.6%; Score 61.6; DB 1; Length 2101;
 Best Local Similarity 70.7%; Pred. No. 8.3e-09;
 Matches 82; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 526 CTGCGATATGCTGATGACGCTTGACGCCACATTCACACATGCTTGGTGAATTC 585
 DB 421 CTATACACAGTCTGATGGAAGATTTATGCTTACATTCACACATTAAGTTTATGTAATTC 480
 OY 586 TTCTGGGACATTCGACGATGACGCTTCACGACATATTCAGAGTCTGCTAGATCG 641
 DB 481 CTCTGATCTGCTCATATATTCCTCCAGGCAATTTGAAGACACATGTTACATTG 536

RESULT 6

Q06086
 ID Q06086 standard; cDNA; 2460 BP.
 AC Q06086;

DE 24-JAN-1991 (first entry)
 DE Plasmid pZPC13 encoding neuronal nicotinic acetylcholine receptor
 DE beta 4 subunit.
 KM Rat; nAChR; ss.

OS Rattus rattus.
 FH Key Location/Qualifiers
 FT cds 61..1548

FT signal_peptide 61..120
 FT /*tag= a
 FT /*tag= b
 FT mat_peptide 121..1545
 FT /*tag= c
 FT /*product=nAChR beta 4 subunit
 FT 416..417
 FT misc_feature 1409..1410
 FT /*tag= d
 FT /label-splice site
 FT 1409..1410
 FT /*tag= e
 FT /label-splice site
 FT 1569..1612
 FT /*tag= f
 FT /number= 1
 FT 1625..1668
 FT /*tag= g
 FT /number= 2
 FT 1683..1726
 FT /*tag= h
 FT /number= 3
 FT 2436..2441
 FT /*tag= i
 FT conflict
 FT /*tag= j
 FT /note= conflict between genomic and cDNA sequences"

FT repeat_unit 1625..1668
 FT repeat_unit 1683..1726
 FT poly_a_signal 2436..2441

replace(268..270,gac)
 /*tag= j
 /note= conflict between genomic and cDNA sequences"

PD W09010648-A.
 PD 20-SEP-1990.
 PD 14-MAR-1990; U01403.
 PR 14-MAR-1989; US-321384.
 PR 12-MAR-1990; US-492555.
 PA (SALK) SALK INST FOR BIOL. STUD.
 PI Heinemann SF. Deneris ES; Duvoisin RM; Patrick JW;
 DR WPI: 90-304987/40.
 DR P-PSDB: R07143.

PT New neuronal nicotinic acetylcholine receptor - compns. contg.
 PT beta 4 sub-unit and DNA sequences encoding them.
 PS Claim 11: Fig 3; 47pp; English.

CC The sequence encodes a novel neuronal nicotinic acetylcholine
 CC receptor subunit, beta 4. Plasmid pZPC13 (ATCC 67893) was
 CC isolated from a cDNA library in lambda ZAP II prep. from rat
 CC mRNA. The sequence of the cDNA in the region encoded by the 5th
 CC exon is identical to that determined for the genomic clone with
 CC the exception of a substn. at posn. 720 of a t for a c (this does
 CC not alter the AA sequence); this may be due to a polymorphism
 CC between the rat strains used for the genomic library and the cDNA
 CC library, or could have resulted from a reverse transcriptase error
 CC during prepn. of the cDNA. A sequence, the core of which is 44 bp
 CC long is repeated three times at the beginning of the 3' UT region.
 CC The function of this remains unknown. The protein subunit can
 CC combine with the known subunits, alpha 2, 3, and -4, and beta 2
 CC to form previously unknown functional receptors. The new subunit
 CC is expressed in the central and peripheral nervous systems and in
 CC PC12 cells.
 CC Sequence 2460 BP; 519 A; 767 C; 587 G; 587 T;

Query Match 7.0%; Score 44.6; DB 1; Length 2460;
 Best Local Similarity 61.7%; Pred. No. 0.0012;
 Matches 71; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 527 TCGCATATGCTGATGACGCTTTGACGCCACATTCACACATGCTTGGTGAATTC 586
 DB 409 TACACAAATGCGGATGACGACCTATGAGGTGCTGCTACACACAGTATGCTGCTCC 468
 OY 587 TCTGGGACATTCGACGATGACGCTTCACGACATATTCAGAGTCTGCTAGATCG 641
 DB 469 AACGACACATTCGACGATGACGCTTCCTCTATCTCAAGAGTCTGCTCAAGATTG 523

RESULT 7
 X53491 standard; DNA; 114955 BP.
 ID X53491;
 AC X53491;
 DT 05-JUL-1999 (first entry)
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 OS Synthetic.
 PN MO913886-A1.
 PD 25-MAR-1999.
 17-SEP-1998; 019419.
 09-JUN-1998; US-093972.
 17-SEP-1997; US-059160.
 (UYEC-) UNIV EAST CAROLINA.
 MYCE JM.
 PI MPI; 99-229400/19.
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 vasoconstriction
 PS Disclosure; Page 37; 120pp; English.
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 directed against at least 2 mRNAs selected from target genes, coding and
 non-coding regions of RNAs corresponding to target genes, gene
 initiation codons, genomic flanking regions, intron-exon borders, the
 5'-end, the 3'-end and the junction between coding and non-coding
 regions and all segments of RNAs encoding proteins associated with one
 or more diseases, conditions or mixtures. The antisense oligonucleotides
 may be derived from sequences X55272-74. These multiple target
 oligonucleotides (specifically X55180-271) can be used for the antisense
 treatment of diseases and conditions. Typical diseases and conditions
 are those associated with impaired respiration and inflammation,
 including lung diseases, pulmonary vasoconstriction, inflammation,
 allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 metastases, as well as all types of cancers which may metastasize or have
 metastasized to the lungs, including breast and prostate cancer.
 CC Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match 6.7%; Score 43; DB 1; Length 114955;
 Best Local Similarity 41.2%; Pred. No. 0.026;
 Matches 73; Conservative 15; Mismatches 89; Indels 0; Gaps 0;

QY 7 GCCACATAGTCCCGCCCACTGCTGCTGCTCCATTTCCAGCCGCTCCACG 66
 ||| :
 DB 105096 GCCGNNNDNCCBGGCGCGCGCGCGCGCGGCGGCGGCGCGCGCG 105155
 ||| :
 QY 67 AGGTACAGCGCGCGGAGAGTGAGAGCGCGAGAGCTGCGCGCGCGCGCTGT 126
 ||| :
 DB 105156 CGGGCCGCGGNNNDNBBGGCGCGCGCGCGCGGCGGCGGCGCGCGG 105215
 ||| :
 QY 127 GCGCGCGCGCGCATGACAGCGCTGCGGACTGCTCTTTCCGCGCCTCCGCG 183
 ||| :
 DB 105216 GGGCGCGGNNNDNNGGGCGCGCGCGCGCGGCGGCGGCGGCGCGG 105272
 ||| :
 RESULT 8
 T95785 standard; CDNA to mRNA; 2383 BP.
 ID T95785;
 AC T95785;
 DT 22-MAY-1998 (first entry)

DE Maize starch synthase type I cDNA
 KW Maize; starch synthase type I; starch; ds.
 OS Zea mays.
 FH Key
 FT CDS
 FT Location/Qualifiers
 2.1951
 /*tag= a
 /*product= starch_synthase_type_I
 DE19619918-A1.
 PD 20-NOV-1997.
 PF 17-MAY-1996; 019918.
 PR 17-MAY-1996; DE-019918.
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 PI Froberg C, Kossman J;
 DR MPI; 98-000821/01.
 P-PDB: W38218.
 PT DNA encoding maize starch synthase type I protein - for producing
 transgenic plants
 PS Claim 1; Pages 16-20; 23pp; German.
 CC The present sequence encodes maize starch synthase type I, useful
 in the production of starch. Starch can be used in various
 conventional starch applications, e.g. starch hydrolysate products,
 CC foods, papermaking, adhesives, textiles, building materials, soil
 CC stabilisation, agrochemicals, pharmaceuticals, cosmetics, coal
 CC briquettes, ore and coal slurries, foundry casting, rubber, leather
 CC and synthetic polymers. The enzyme produces a starch starched to have
 CC different physicochemical properties, especially viscosity and
 CC gelling properties, from wild type starch.
 SQ Sequence 2383 BP; 582 A; 546 C; 633 G; 622 T;

Query Match 6.7%; Score 42.8; DB 1; Length 2383;
 Best Local Similarity 55.3%; Pred. No. 0.0039;
 Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 58 GCTCCACAGAGGTACAGCGCGCGGAGAGTGGAGCCGACAGCTCGCGCGGCC 117
 ||| :
 DB 121 GCTCCAGCGCGGTGCTGCGCGCGCGGTGCTGCGGAGCTGACAGAGGCGCCGCGCC 180
 ||| :
 QY 118 CGGCTGGGCGCGCGGCATACAGCGGCTGGGAGTGGCTCTTTCCGCGCCCTCC 177
 ||| :
 DB 181 GCGCCGATGACACCGCGGCTGCTGCGCGCCCGCTGCGCGCTCTCTGCGCGCC 240
 ||| :
 QY 178 CGCGAGGTGAGGAGATGTCATGTC 207
 ||| :
 DB 241 GCGCGAGCCACAGGCTGAGCGCGCATTGC 270
 ||| :
 RESULT 9
 X28103/c
 ID X28103 standard; DNA; 2106 BP.
 AC X28103;
 DT 11-JUN-1999 (first entry)
 DE Frec11 gene.
 KW Frec11; transcription factor; adipose tissue expressed gene; obesity;
 KW lipid metabolism; adipocyte differentiation; obesity-related condition;
 KW non-insulin dependent diabetes mellitus; cardiovascular disease;
 KW metabolic condition; anorexia; bulimia; therapy; ss.
 OS Homo sapiens.
 PN MO9854216-A1.
 PD 03-DEC-1998.
 PF 26-MAY-1998; SE0989.
 PR 26-MAY-1997; SE-001963.
 PA (PHMA) PHARMACIA & UPJOHN AB.
 PI Carlsson P, Enerbaeck S;
 DR MPI; 98-070211/06.
 P-PDB: Y01097.
 PT New transcription factor designated FREC11 - which regulates
 PT adipose tissue expressed genes involved in lipid metabolism and
 PT adipocyte differentiation, used to, e.g. inhibit gene expression
 PS Claim 1; Fig 1; 23pp; English.
 CC This sequence represents the Frec11 gene of the invention. The encoded
 CC protein has transcriptional regulatory function directed against adipose
 CC tissue expressed genes. The protein is involved in lipid metabolism


```

FT cds 87. 1583
PD WO9641876-A1. /tag= a
PF 27-DEC-1996.
PR 07-JUN-1996; 009775.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Elliott KJ, Harpold MM;
DR WPI: 97-065463/06.
DR P-PSDB: W09027.
PT Nucleic acids encoding nicotinic acetylcholine receptor sub-units -
PS used in screening to determine the effect of drugs on the receptor
PD disclosure: Page 83-85; 108pp; English.
CC A DNA sequence (T48241) codes for the beta-4 subunit (W09027) of
CC the human neuronal nicotinic acetylcholine receptor (nAChR). Host
CC cells, esp. mammalian cells or amphibian oocytes, carrying beta-4
CC nucleic acids, opt. in combination with other alpha and/or beta
CC subunit nucleic acids (see also T48232-40), express recombinant
CC nAChR subunits useful for identifying cpds. that modulate the
CC activity of human nAChRs.
CC Sequence 1915 BP; 374 A; 631 C; 505 G; 405 T;

Query Match 5.7%; Score 36.6; DB 1; Length 1915;
Best Local Similarity 57.4%; Pred. No. 0.26;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 527 TCGCATTTGCTGATGAGCGCTTTCAGCCACATTCACACTACGTTGTAATCT 586
DB 438 TACACACAGCGCGGACGGAGCTATGAGTGTCTGTCTACACCACTGATAGCGGTCC 497
OY 587 TCTGGCATTGCCAGTACCTGCTCCAGGACATTCAGAGTTCTCTGCTACATCG 641
DB 498 AACGCGAGCGCTCTGTGCTGCGCCCTCCATCTACAGAGCGCTCGAAGATTG 552

RESULT 13
V68408/c
ID V68408 standard; cDNA to mRNA; 5561 BP.
AC V68408;
DE 05-MAY-1999 (first entry)
DE Human BAZ1-beta cDNA #1
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 346..4929
/Tag= a "BAZ1-beta"
/product= "BAZ1-beta"

PD WO9847920-A1.
PD 29-OCT-1998.
PF 17-APR-1998; J01783.
PR 24-OCT-1997; JP-310027.
PR 18-APR-1997; JP-116570.
PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
DR WPI: 98-583603/49.
DR P-PSDB: W81172.
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PT and other proliferative disorders
PS Claim 2; Page 125-137; 187pp; Japanese.
CC This sequence encodes the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ;
CC Bromodomain with Atypical zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).
CC Sequence 5561 BP; 1665 A; 1261 C; 1471 G; 1164 T;

```

```

Query Match 5.7%; Score 36.4; DB 1; Length 5561;
Best Local Similarity 58.2%; Pred. No. 0.51;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 6 CGCCACATAGCTCCCGCCAGAGTCGTCGCTCCCTTCGATTTTCACGCCGCTCCAC 65
DB 156 CGGCGACAGAGCTCCCGCCACACCGCGCTCCAGACAGCCCCCGCCAGCTCCGCT 97
OY 66 GAGGCTCAGCGCGCGGAGAGTGAGCCCGGAGAGCTCGCGCGGGG 115
DB 96 TCGGCTCCCGGCGCGGAGAGTACAGACTCTCTCAGACAGACCGGAGG 47

RESULT 14
V68409/c
ID V68409 standard; cDNA to mRNA; 5573 BP.
AC V68409;
DE 05-MAY-1999 (first entry)
DE Human BAZ1-beta cDNA #2.
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 346..4941
/Tag= a
/product= "BAZ1-beta"

PD WO9847920-A1.
PD 29-OCT-1998.
PF 17-APR-1998; J01783.
PR 24-OCT-1997; JP-310027.
PR 18-APR-1997; JP-116570.
PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
DR WPI: 98-583603/49.
DR P-PSDB: W81173.
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PT and other proliferative disorders
PS Claim 2; Page 145-157; 187pp; Japanese.
CC This sequence encodes the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ;
CC Bromodomain with Atypical zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).
CC Sequence 5573 BP; 1668 A; 1267 C; 1472 G; 1166 T;

Query Match 5.7%; Score 36.4; DB 1; Length 5573;
Best Local Similarity 58.2%; Pred. No. 0.52;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 6 CGCCACATAGCTCCCGCCAGAGTCGTCGCTCCCTTCGATTTTCACGCCGCTCCAC 65
DB 156 CGGCGACAGAGCTCCCGCCACACCGCGCTCCAGACAGCCCCCGCCAGCTCCGCT 97
OY 66 GAGGCTCAGCGCGCGGAGAGTGAGCCCGGAGAGCTCGCGCGGGG 115
DB 96 TCGGCTCCCGGCGCGGAGAGTACAGACTCTCTCAGACAGACCGGAGG 47

RESULT 15
O84888/c
ID O84888 standard; cDNA to mRNA; 1942 BP.
AC O84888;
DR 26-OCT-1995 (first entry)
DE DNA encoding cytoplasmic tyrosine kinase.

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 22:05:19 ; Search time 1962.17 Seconds
(without alignments)
1038.941 Million cell updates/sec

Title: US-08-956-518A-103

Perfect score: 641
Sequence: 1 CAGCCGCCACATAGCTCCC.....CAAGAGTCTCTGCTACATCG 641

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

ibase :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_v1:*
18: em_fun:*
19: em_hlg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_hlg1:*
35: gb_hlg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	641	100.0	641	11	AF029838	AF029838 Homo sapi

2	503	78.5	689	11	AF029839	AF029839 Homo sapi
3	217.2	33.9	291	9	HS2E10F	Z58126 H.sapiens C
4	196.6	30.7	137699	11	AC005630	AC005630 Homo sapi
5	194.2	30.3	298	9	HS2G7F	Z60646 H.sapiens C
6	192	30.0	1712	11	AF036903	AF036903 Homo sapi
7	171.8	26.8	192439	35	AC007562	AC007562 Homo sapi
8	163.4	25.5	178307	35	AC006328	AC006328 Homo sapi
9	156.8	24.5	1836	11	AF037646	AF037646 Homo sapi
10	108	16.8	1559	9	HUMA7A	L25827 Human a7 n1
11	108	16.8	2087	10	HSARA7A	X70297 H.sapiens m
12	108	16.8	1509	10	HSNAC7A7	Y08420 H.sapiens m
13	108	16.8	1876	10	HSN62436	U62436 Human nicot
14	106.4	16.6	1555	10	HSCHRNA7A	U23141 H.sapiens C
15	106.4	16.6	1977	10	HSU040583	U40583 Human alpha
16	92.6	14.4	5281	9	HSU08191	U08191 Human R Kap
17	92	14.4	2106	12	MUSNARS	L37663 Mus musculus
18	90.4	14.1	3030	12	RATNARAD	L31619 Rattus ratt
19	88.8	13.9	1551	3	BRAT7	S53987 nicotinic r
20	88.8	13.9	1551	3	BRAT7	X93604 B.taurus nr
21	74.4	11.6	2696	4	GGATNAREC	X52285 Chicken alp
22	74.4	11.6	2037	4	GGNNA7A7A	X68586 G.gallus mr
23	62.4	9.7	171	9	HS165HIF	Z57156 H.sapiens C
24	61.8	9.6	180	9	HS115G2F	Z54415 H.sapiens C
25	61.6	9.6	2090	4	GGARNAREC	X52296 Chicken alp
26	60.4	9.4	189	9	HS31DBR	Z61332 H.sapiens alp
27	58.2	9.1	3029	37	AF143847	AF143847 Heliothis
28	50	7.8	89871	37	AC005890	AC005890 Drosophila
29	48.2	7.5	32039	1	SC7A1	AL034447 streptomy
30	46.4	7.2	336	14	G47584	G47584 Z25357.1 Ze
31	46.2	7.2	76947	36	AC004326	AC004326 Drosophila
32	46.2	7.2	3629	37	AF143846	AF143846 Heliothis
33	44.6	7.0	2493	12	RNNIACETE	X15834 R.norvegicus
34	44.6	7.0	2461	12	RNU42976	U42976 Rattus norv
35	44.6	6.9	1008	12	RATNACHRF5	M33952 Rat neurona
36	42.8	6.7	143664	35	AC007291	AC007291 Drosophila
37	42.4	6.6	3459	10	HSAR12582	AJ012582 Homo sapi
38	42.4	6.6	785	10	HSAR13727	AJ13727 Homo sapi
39	42.2	6.6	107889	11	AC004519	AC004519 Homo sapi
40	41.8	6.5	1574	11	AF069682	AF069682 Homo sapi
41	41.6	6.5	120766	42	AC004150	AC004150 Homo sapi
42	41.2	6.4	2672	3	RABTHRP	T05603 Rabbit mult
43	41.2	6.4	4085	4	CHKACHR02	K02904 Chicken nic
44	41.2	6.4	3289	10	HSMEF1	Y08223 H.sapiens M
45	41.2	6.4	173422	35	AC007041	AC007041 Homo sapi

ALIGNMENTS

RESULT 1
AF029838 641 bp mRNA PRI 16-DEC-1998
DEFINITION Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
ACCESSION AF029838
NID 93757793
VERSION AF029838.1 GI:3757793
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 641)
Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,
Buiting,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.,
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)
JOURNAL 99000837
MEDLINE
REFERENCE
2 (bases 1 to 641)
Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
Breese,C., Davis,A., Hopkins,J. and Freedman,R.
Direct Submission
AUTHORS
TITLE

JOURNAL Submitted (15-OCT-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA

FEATURES Location/Qualifiers

source 1..641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR1"
/note="RACE product A/B/C/D"
1..>641
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

misc-feature
BASE COUNT 151 a 192 c 163 g 135 t
ORIGIN

Query Match 100.0%; Score 641; DB 11; Length 641;
Best Local Similarity 100.0%; Pred. No. 3.4e-152;
atches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CAGCGCCGACATAGCTCCCGCCAAATCTCGGTGCGCCCTTGCCATTTCACGCCCGCT 60
   |||||||
Db 1 CAGCGCCGACATAGCTCCCGCCAAATCTCGGTGCGCCCTTGCCATTTCACGCCCGCT 60
QY 61 CCCACAGAGGTTCAGCGCGGGGAGAGTGAAGCCGCGAGAGCTGGCGGGGGCCCG 120
   |||||||
Db 61 CCCACAGAGGTTCAGCGCGGGGAGAGTGAAGCCGCGAGAGCTGGCGGGGGCCCG 120
QY 121 CCTGTGGCGCGCGCCATGACAGCGGCTCGGAGCTGCTCTTTCCGCGCCCTCCCG 180
   |||||||
Db 121 CCTGTGGCGCGCGCCATGACAGCGGCTCGGAGCTGCTCTTTCCGCGCCCTCCCG 180
QY 181 CGGAGGTGAGGGGAAATATGCTCATGTCAAGGTTCAAGGCCAAACCAATTACTGCTC 240
   |||||||
Db 181 CGGAGGTGAGGGGAAATATGCTCATGTCAAGGTTCAAGGCCAAACCAATTACTGCTC 240
QY 241 TATCTTCCAGAGAACCCAGAGCCAGCGCGGCTCAGCGCCCAACCGAATTAAAGAT 300
   |||||||
Db 241 TATCTTCCAGAGAACCCAGAGCCAGCGCGGCTCAGCGCCCAACCGAATTAAAGAT 300
QY 301 TACAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 360
   |||||||
Db 301 TACAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 360
QY 361 GACTCAGATCTTGTTCACCCCATTTATGACATCAAGAGTGCAGAAAGCACTCTGA 420
   |||||||
Db 361 GACTCAGATCTTGTTCACCCCATTTATGACATCAAGAGTGCAGAAAGCACTCTGA 420
QY 421 CAATATGAAACACACCATGCGTTAAATTGATGCAAAAATATGATGATACACGA 480
   |||||||
Db 421 CAATATGAAACACACCATGCGTTAAATTGATGCAAAAATATGATGATACACGA 480
QY 481 TTTTCAGTTCATTCATTCAGCATTGTGATAGCTGCAAAATGCGATTGCTGA 540
   |||||||
Db 481 TTTTCAGTTCATTCATTCAGCATTGTGATAGCTGCAAAATGCGATTGCTGA 540
QY 541 TGAGCCCTTTGACGCCACATTCACACTAAGCTGTGGTAATCTCTTGCGGCAATTGCCA 600
   |||||||
Db 541 TGAGCCCTTTGACGCCACATTCACACTAAGCTGTGGTAATCTCTTGCGGCAATTGCCA 600
QY 601 GATCTGCTCCAGCATATTCAGAGTTCTGCTACATCG 641
   |||||||
Db 601 GATCTGCTCCAGCATATTCAGAGTTCTGCTACATCG 641

```

RESULT 2
AF029839 689 bp mRNA PRI 16-DEC-1998
LOCUS Homo sapiens alpha 7 neuronal nicotinic receptor mRNA sequence.
DEFINITION
ACCESSION AF029839
NID 93757794
VERSION AF029839.1 GI:3757794

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 689)
Gault,J., Robinson,M., Berger,R., Drebing,C., Logel,J., Hopkins,J.,
Moore,T., Jacobs,S., Meriwether,J., Choi,M.J., Kim,E.J., Walton,K.,
Bulling,K., Davis,A., Breese,C., Freedman,R. and Leonard,S.
Genomic organization and partial duplication of the human alpha7
neuronal nicotinic acetylcholine receptor gene
Genomics 52 (2), 173-185 (1998)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 689)
Leonard,S., Gault,J., Logel,J., Drebing,C., Robinson,M., Berger,R.,
Breese,C., Davis,A., Hopkins,J. and Freedman,R.
Direct Submission
Submitted (15-OCT-1997) Psychiatry, University of Colorado Health
Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
Location/Qualifiers

FEATURES
source 1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q14"
/clone="CHRNA7-DR2"
/note="RACE product A/C/D"
1..>689
/note="alpha 7 neuronal nicotinic receptor mRNA; used in alternative splicing"

BASE COUNT 142 a 210 c 196 g 141 t
ORIGIN

Query Match 78.5%; Score 503; DB 11; Length 689;
Best Local Similarity 90.0%; Pred. No. 2.5e-117;
Matches 577; Conservative 0; Mismatches 0; Indels 64; Gaps 1;

```

QY 1 CAGCGCCGACATAGCTCCCGCCAAATCTCGGTGCGCCCTTGCCATTTCACGCCCGCT 60
   |||||||
Db 113 CAGCGCCGACATAGCTCCCGCCAAATCTCGGTGCGCCCTTGCCATTTCACGCCCGCT 172
QY 61 CCCACAGAGGTTCAGCGCGGGGAGAGTGAAGCCGCGAGAGCTGGCGGGGGCCCG 120
   |||||||
Db 173 CCCACAGAGGTTCAGCGCGGGGAGAGTGAAGCCGCGAGAGCTGGCGGGGGCCCG 232
QY 121 CCTGTGGCGCGCGCCATGACAGCGGCTCGGAGCTGCTCTTTCCGCGCCCTCCCG 180
   |||||||
Db 233 CCTGTGGCGCGCGCCATGACAGCGGCTCGGAGCTGCTCTTTCCGCGCCCTCCCG 292
QY 241 TATCTTCCAGAGAACCCAGAGCCAGCGCGGCTCAGCGCCCAACCGAATTAAAGAT 300
   |||||||
Db 241 TATCTTCCAGAGAACCCAGAGCCAGCGCGGCTCAGCGCCCAACCGAATTAAAGAT 300
QY 293 CGGAGGTGAGGGGAAAGATGCTCATGTCAAGGTTCAAGGCCAAACCGAATTACTGCGCTC 352
   |||||||
Db 293 CGGAGGTGAGGGGAAAGATGCTCATGTCAAGGTTCAAGGCCAAACCGAATTACTGCGCTC 352
QY 301 TACAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 360
   |||||||
Db 301 TACAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 360
QY 413 TACAAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 472
   |||||||
Db 413 TACAAGTGACACCTGATCAGCAGGACCTGGAATCCAGATGAGAGCTTATCTACAC 472
QY 473 GACTCAGATCTTGTTCACCCCATTTATGACATCAAGAGTGCAGAAAGCACTCTGA 532
   |||||||
Db 473 GACTCAGATCTTGTTCACCCCATTTATGACATCAAGAGTGCAGAAAGCACTCTGA 532
QY 421 CAATATGAAACACACCATGCGTTAAATTGATGCAAAAATATGATGATACACGA 480
   |||
Db 533 CAA-----TTCCAAATGCAATTCAGCAGCTTGTGATAGCTCAAACTCGATATTGCTGA 540
   |||||||
QY 481 TTTTCAGTTCATTCATTCAGCATTGTGATAGCTGCAAAATGCGATTGCTGA 540
   |||||||
Db 536 -----TTCCAAATGCAATTCAGCAGCTTGTGATAGCTCAAACTCGATATTGCTGA 588

```


QY 541 TGAGCGCTTGAGCCACATTCACACTAAGCGTGGTGAATTCCTTGGGATTCGCA 600
|||||
Db 589 TGAGCGCTTGAGCCACATTCACACTAAGCGTGGTGAATTCCTTGGGATTCGCA 648
QY 601 GTACTGCTCCAGGATATTCAGAGTTCCTGCTACATCG 641
|||||
Db 649 GTACTGCTCCAGGATATTCAGAGTTCCTGCTACATCG 689
RESULT 3
HS2E10F 291 bp DNA PRI 18-OCT-1995
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 2e10, forward
DEFINITION read cpg2e10.f.titld.
ACCESSION 258126
NID 91029357
VERSION 258126.1 GI:1029357
KEYWORDS Cpg Island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS MacDonald, M., Huckle, E., Wilkinson, P. and Mickle, G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humuery@sanger.ac.uk
COMMENT 2 (bases 1 to 291)
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
Purification of Cpg Islands using a methylated DNA binding column
Nat. Genet. 6 (3), 236-244 (1994)
94282070
Vector: pGEM-52f(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hmp.mrc.ac.uk/ for details
or Contact: biochem@hmp.mrc.ac.uk.
Location/Qualifiers
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_id="CGI-1"
/clone="2e10"
BASE COUNT 44 a 95 c 103 g 45 t 4 others
TGIN
FEATURES
source
Query Match 33.9%; Score 217.2; DB 9; Length 291;
Best Local Similarity 97.3%; Pred. No. 3.5e-45;
Matches 219; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

AC005630
LOCUS AC005630 137699 bp DNA PRI 15-JAN-1999
DEFINITION Homo sapiens clone DJ1129D05, complete sequence.
ACCESSION AC005630
NID 94159882
VERSION AC005630.1 GI:4159882
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 137699)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (bases 1 to 137699)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 137699)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jan 15, 1999 this sequence version replaced gi:3550965.
Location/Qualifiers
1..137699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1129D05"
BASE COUNT 36908 a 30670 c 32412 g 37709 t
ORIGIN
FEATURES
source
Query Match 30.7%; Score 196.6; DB 11; Length 137699;
Best Local Similarity 80.2%; Pred. No. 1.2e-39;
Matches 243; Conservative 0; Mismatches 59; Indels 1; Gaps 1;
QY 1 CAGGCGCCACATAGCT-CCGCGCAAGTCCGCGGCGCCCTGCAATTTCCAGCGCGC 59
|||||
Db 47458 CAGGCGCCACATAGCTCCCGCAAGTCCGCGGCGCCCTGCAATTTCCAGCGCGC 47517
QY 60 TCCACAGAGGTTCACGCGCGCGGAGAGGTGAGAGCCGCGAGAGTCCGCGCGCGCGC 119
|||||
Db 47518 TCCGACTAGTGAAGGATTCAGAGAGGAGAGGAGAGTGGAGAGCGCCCGCAGAGTCC 47577
QY 120 GCCTGGTGGCGCGGCGCATGACAGCGGCTCGGAGCTGCTCTTTCCGCGCGCTCCG 179
|||||
Db 47578 CCGGCTGGTTCGCGCGCGGAGAGCGGCTCCGAGCGGCTCACCTTCCGCGCGCTCCG 47637
QY 180 CCGAGGTGAGGGGAGAGTCCATGCTCAGGCTCAAGGCCAAGCAAGTACAGCGCT 239
|||||
Db 47638 CCGAGGTGAGGGGAGAGTCCATGCTCAGGCTCAAGGCCAAGTACAGCGCT 47697
QY 240 CTATCTTCAGAGAGAACAGAGACAGAGCCGCGCTCAGCGCCACCGCAATTAAGA 299
|||||
Db 47698 CTATCTTCAGAGAGAACAGAGAGAGAGCGCGCGCGCAGCTCAGCTCAGCAATCAGCG 47757
QY 300 TTA 302
Db 47758 TGA 47760
RESULT 5
HS2G7F 298 bp DNA PRI 19-OCT-1995
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 2g7, forward
DEFINITION read cpg2g7.f.tla.
ACCESSION Z60646
NID 91032750

VERSION 260646.1 GI:1032750
 KEYWORDS CPG island: genomic MseI fragment.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 298)
 TITLE Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
 JOURNAL Direct Submission
 Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
 2 (bases 1 to 298)
 REFERENCE Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
 AUTHORS Purification of CPG islands using a methylated DNA binding column
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
 MEDLINE 94282070
 COMMENT Vector: pGEM-5zf(-)
 Clones are available from the UK MRC Human Genome Mapping Project
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
 http://www.hgmp.mrc.ac.uk/ for details
 or contact: biolhelp@hgmp.mrc.ac.uk.
 Location/Qualifiers
 1. 298
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /sex="male"
 /dev_stage="adult"
 /tissue_type="blood"
 /clone_lib="CGI-1"
 /clone="297"

BASE COUNT 47 a 99 c 102 g 48 t 2 others
 ORIGIN

Query Match 30.3%; Score 194.2; DB 9; Length 298;
 Best Local Similarity 94.5%; Pred. No. 2, 3e-39;
 Matches 223; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

QY 4 GCCGCACATAGTCTCCGCGCAAGTCTCGGTGCGCCCTTGCATTTTCACGCGCGCTCC 63
 DB 67 GCCGCACATAGTCTCCGCGCAAGTCTCGGTGCGCCCTTGCATTTTCACGCGCGCTCC 126

QY 64 ACAGAGTCAAGCGCGCGGAGAGAGTGGAGCGCGGAGAGAGTGGCGCGCGCGCGCT 123
 DB 127 ACAGAGTCAAGCGCGCGGAGAGAGTGGAGCGCGGAGAGAGTGGCGCGCGCGCT 183

QY 124 GGTGGCGCGCGCGCGCGGAGAGAGTGGAGCGCGCGCGCGCGCGCGCT 183
 DB 184 GGTGGCGCGCGCGCGCGGAGAGAGTGGAGCGCGCGCGCGCGCGCGCT 243

QY 184 AGGTGAGGAGAGATGTCATGTCAGAGGTTCAGAGCGCAACCGAGTACTGGCCT 239
 DB 244 A-GTGAGGGAGAGATGTCATGTCAGAGGTTCAGAGCGCAACCGAGTACTGGCCT 298

RESULT 6
 AF036903 1712 bp mRNA PRI 16-OCT-1998
 LOCUS Homo sapiens alpha-7 neuronal nicotinic acetylcholine receptor
 DEFINITION
 AF036903
 ACCESSION
 NID 93757807
 VERSION 3757807
 KEYWORDS AF036903.1 GI:3757807
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 1712)
 REFERENCE Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J.,
 Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K.,
 Bulting, K., Davis, A., Breese, C., Freedman, R. and Leonard, S.
 COMMENT Genomic organization and partial duplication of the human alpha7

JOURNAL neuronal nicotinic acetylcholine receptor gene (CHRNA7)
 MEDLINE Genomics 52 (2), 173-185 (1998)
 REFERENCE 99000837
 2 (bases 1 to 1712)
 TITLE Leonard, S., Gault, J., Logel, J., Drebing, C., Robinson, M., Berger, R.,
 JOURNAL Breese, C., Davis, A., Hopkins, J. and Freedman, R.
 Submitted (03-DEC-1997) Psychiatry, University of Colorado Health
 Sciences Center, 4200 E. 9th Ave., Denver, CO 80262, USA
 1. 1712
 Location/Qualifiers
 1. 1712
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /db_xref="dbEST:W52861"
 /db_xref="dbEST:AA037389"
 /chromosome="15"
 /clone="321242"
 /clone_lib="Soares papathroid tumor NBHRA library,
 Reseach Genetics/IMAGE Consortium, LBNL"
 <1. 1712
 /note="alpha-7 neuronal nicotinic acetylcholine receptor;
 alternatively spliced"

BASE COUNT 356 a 489 c 457 g 410 t
 ORIGIN

Query Match 30.0%; Score 192; DB 11; Length 1712;
 Best Local Similarity 100.0%; Pred. No. 1e-38;
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 ATTGATGCAAAATATTCATCTACACAGATTTTCAGTTCAGTTCATTCAGCAGATT 509
 DB 10 ATTGATGCAAAATATTCATCTACACAGATTTTCAGTTCAGTTCATTCAGCAGATT 69

QY 510 TGTGATAGCTCAAAACGCGATATTCGTGATGAGCGCTTACGCCATATTCACAGATT 569
 DB 70 TGTGATAGCTCAAAACGCGATATTCGTGATGAGCGCTTACGCCATATTCACAGATT 129

QY 570 ACGTGTGATGATCTTCTGCGCATTCGCGCATTCCTGCTCCAGGCAATTCAGAGATT 629
 DB 130 ACGTGTGATGATCTTCTGCGCATTCGCGCATTCCTGCTCCAGGCAATTCAGAGATT 189

QY 630 CCTGCTACATCG 641
 DB 190 CCTGCTACATCG 201

RESULT 7
 AC007562 192439 bp DNA HTG 05-JUN-1999
 LOCUS Homo sapiens clone NH0497C14, WORKING DRAFT SEQUENCE, 4 unordered
 DEFINITION
 AC007562
 ACCESSION
 NID 95001485
 VERSION AC007562.2 GI:5001485
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 192439)
 REFERENCE Waterston, R.H.
 JOURNAL The sequence of Homo sapiens clone
 unpublished
 2 (bases 1 to 192439)
 TITLE Waterston, R.H.
 JOURNAL Direct Submission
 Submitted (14-MAY-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Jun 5, 1999 this sequence version replaced g1:4827309.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces

	MEDLINE	99000837
	REFERENCE	2 (bases 1 to 1836)
	AUTHORS	Leonard,S., Gauld,J., Logel,J., Drebing,C., Robinson,M., Berger,R., Biese,C., Davis,A., Hopkins,J. and Freedman,R.
	TITLE	Direct Submission
	JOURNAL	Submitted (03-DEC-1997) Psychiatry, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Denver, CO 80202, USA
	FEATURES	Location/Qualifiers
	SOURCE	1..1836
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/db_xref="dbEST:F03952"
		/db_xref="dbEST:N73891"
		/chromosome="15"
		/clone="297134"
		/clone_lib="Scares fetal liver spleen library, Research Genetics/IMAGE Consortium, LBNL"
		/note="unprocessed mRNA with intron"
		<1..1836
		/note="alpha-7 neuronal nicotinic acetylcholine receptor precursor; Intron included in 5' region"
	misc_feature	
	BASE COUNT	383 a 503 c 469 g 481 t
	ORIGIN	
	Query Match	24.5% Score 156.8; DB 11; Length 1836;
	Best Local Similarity	98.8%; Pred. No. 7.9e-30;
	Matches 158; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
OY	482 TTTGAGTTCATTGCTATCCAGATTGTGGATGCACAACTGGCATTTGGTGAT	541
DB	164 TTCTGATTCATTCTCTAATCCAGATTGTGGATGCACAACTGGCATTTGGTGAT	223
OY	542 GAGCGCTTTGAGCGCACATTCACACTAACGTTGGTGATTCCTTGGCATTGCCAG	601
DB	224 GAGCGCTTTGAGCGCACATTCACACTAACGTTGGTGATTCCTTGGCATTGCCAG	283
OY	602 TACCTGCTCCAGGCATATTCGAAGATTCTCTACATCG	641
DB	284 TACCTGCTCCAGGCATATTCGAAGATTCTCTACATCG	323
RESULT 10	HOMANAR	
LOCUS	HUWA7NAR	1559 bp mRNA PRI 18-AUG-1994
DEFINITION	Human a7 nicotinic acetylcholine receptor mRNA.	
SSION	L25827	
VERSION	G438616	
KEYWORDS	L25827.1 GI:438616	
SOURCE	nicotinic acetylcholine receptor alpha 7 subunit; transmembrane protein.	
ORGANISM	Homo sapiens (library: Clontech HL1065b; ATCC 37433) fetus brain CDNA to mRNa.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Dovelette-Stamm,L., Monteggia,L.M., Donnelly-Roberts,D., Wang,M.T., Lee,J., Tian,J. and Giordano,T.	
TITLE	Cloning and sequence of the human a7 nicotinic acetylcholine receptor	
JOURNAL	Drug Dev. Res. (1993). In press	
FEATURES	Location/Qualifiers	
SOURCE	1..1559	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/db_xref="taxon:9606"	
	/dev_stage="fetus"	
	/tissue_type="brain"	
	/tissue_lib="Clontech HL1065b; ATCC 37433"	
BASE COUNT	314 a 463 c 440 g 342 t	
ORIGIN		

mat_peptide 441 a 601 c 573 g 472 t

BASIS COUNT 441 a 601 c 573 g 472 t

RRIGIN

Query Match 16.8% Score 108; DB 9; Length 1559;
Best Local Similarity 95.7%; Pred. No. 1.7e-17;
Matches 111; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DQ 526 CTGGCATATTCGTGATGAGCGCTTTGACGCCACATTCACACTAAGTGTGGTAATC 585
||| |
DQ 319 CTAATAACAGTGCTGATGAGCGCTTTGACGCCACATTCACACTAAGTGTGGTAATC 378
|||||
QY 586 TTTCGGGCATTGCCAGTACCTGCTCCACAGGCATATTCAGAAGTTCCTCATCG 641
|||||
Db 379 TTCGGGCATTGCCAGTACCTGCTCCACAGGCATATTCAGAAGTTCCTCATCATCG 434
|||||

RESULT 11

HSARA7A
LOCUS
DEFINITION
H.sapiens mRNA for neuronal nicotinic acetylcholine receptor alpha-7 subunit.
ACCESSION
X70297
NID
9496606
VERSION
X70297.1 GI:496606
KEYWORDS
neuronal nicotinic acetylcholine receptor alpha-7 subunit.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2087)
Peng,X., Katz,M., Gerzanich,V., Anand,R. and Lindstrom,J.
Human alpha 7 acetylcholine receptor: cloning of the alpha 7
subunit from the SH-SY5Y cell line and determination of
pharmacological properties of native receptors and functional alpha
7 homomers expressed in Xenopus oocytes
Mol. Pharmacol. 45 (3), 546-554 (1994)
MEDLINE
94195283
REFERENCE
2 (bases 1 to 2087)
Katz,M.
TITLE
Direct Submission
JOURNAL
Submitted (04-FEB-1993) M. Katz, Univ of Pennsylvania School of
Medicine, Dept of Neuroscience, 36th & Hamilton walk, Rm 335
Stemmer Hall, Philadelphia, PA 19104, USA
FEATURES
source
1. 2087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/cell_type="neuroblastoma"
/cell_line="SHSY-5Y"
/clone_id="lambda zap II"
/clone="SHSY3"
104..175
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
CODON_start=1
/codon_start=1
/product="neuronal nicotinic acetylcholine receptor
alpha-7 subunit"
protein_id="CAA49778.1"
/db_xref="PIR:P496607"
/db_xref="GI:496607"
/db_xref="SWISS-PROT:P36544"
translation="MRCSPGVWGLAAALSHVSLOGEPORRLYLKELKNYNPLERPV
ANDSQLTYFSLIOLIMDVDERKNVLNTNLQMSMTDHPDMSEVPGRKYTF
PDSQIKRPDLILNSADERDATFRHNLYLVNSSGHCQYLPDGIFKSCYIDVRFPED
VDFCKTKFGSWISGWSLDIOMQEADISGITPNGEWDLVGIPGRSEREFECCKEPP
DYFTLTVRRRTLYIGLINTLIPCULSALTIPADSGEKISLGITVLDSLITYEM
LLVAELIMPATSDEVPFLIAOYFASTMITIGSVVVIVTIYLOHHHDPDGSKPMKTRVI
LTLMWCFLMKRPGEDEKRVAPACOHKORSLSAVESMAVGPAPSNGNLITGPFGL
DYGHCVPTDSDGVYGARMACSPRHDELHLHGQDPGPDPLAKILEEVRIANRFQC
DSEBANVCSEKRPAACYVDRLCLMAFSVFITICTIGILMSAPNFVEAVSKDFA"
176..1603

3'UTR 1582 1876 531 g 423 t
BASE COUNT 369 a 553 c
ORIGIN

Query Match 16.8%; Score 108; DB 10; Length 1876;
Best Local Similarity 95.7%; Pred. No. 1.7e-17;
Matches 110; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 526 CTGCGATATTCGTAGTACGCGCTTTGACGCGACATTCACACTACGTTGGTGAATTC 585
||| |
Db 414 CTAAACAGTGTGATGAGCGCTTTGACGCGACATTCACACTACGTTGGTGAATTC 473
||| |
QY 586 TTCTGGGCGATTCGACGAGTCTGCTCCAGGCGATATTCAGAGTCTGCTACATCG 641
||| |
Db 474 TTCTGGGCGATTCGACGAGTCTGCTCCAGGCGATATTCAGAGTCTGCTACATCG 529

LOCUS 1555 bp mRNA PRI 03-MAR-1994
DEFINITION H.sapiens CHRNA7 mRNA, 3' end.
ACCESSION 223141
NID 6457736
VERSION 223141.1 GI:457736
KEYWORDS alpha/ nicotinic receptor subunit; cholinergic receptor; CHRNA7 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Chini, B., Raimond, E., Eljoughen, A.B., Morali, D., Balzaretto, M. and Heinemann, S.
TITLE Molecular cloning and chromosomal localization of the human alpha 7-nicotinic receptor subunit gene (CHRNA7)
JOURNAL Genomics 19 (2), 379-381 (1994)
MEDLINE 94245214
REFERENCE 2 (bases 1 to 1555)
AUTHORS Chini, B.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1993) BICE CHINI, CCIPE, Rue de la Cardonille, Montpellier, 34094 Cedex 5, France
FEATURES
source
1. 1555
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="retina"
/clone_lib="lambda gt11"
<1.1437
/gene="CHRNA7"
/product="cholinergic receptor, nicotinic, alpha polypeptide 7"
1. 1440
/gene="CHRNA7"
<1.1440
/codon_start=1
/product="cholinergic receptor, nicotinic, alpha polypeptide 7"
/protein_id="CA80672.1"
/db_xref="PIR:G457737"
/db_xref="GI:457737"
/translation="EFQRLKYLKLVNPLERPVANDSOPLVFSLIQLQIMDVDE
KNQVLTINIMLQMSWTDLQNVSEYPGVKTYRPPDGOIMKPDILLVNSADERDPA
FTNVLVNSGHCQYLPQGIFFKSCYIDRWPFQVQHCRIKGSWYSGMSLDQMO
EADISGVIENGENDIVIGIKRSERYECCKEPPYPTTVTRRTLTIGLNLIP
VLISALVFLPADSGEKISLITVLSTVFMLVAEIMPATSDVPLIAQFAS
TMIVGLSVYVTVIVLQHHHDPDGMKPKMTVILLNMCAMFLRKMRGEGEDKVRPAC
OHKORSSSLASVSAVAPPASNGNLIVIGFGDGVHCVPTPDGGVCGAMPAPT
HDEHLHGQGPPEGDPLAKILEEVRVIANRFRCQDESAVQSEMKFACVVDRLCLM
AFSVFTICTIGILMSAPNFVEAVSKDFA"

BASE COUNT 320 a 454 c 434 g 347 t
ORIGIN

Query Match 16.6%; Score 106.4; DB 10; Length 1555;
Best Local Similarity 94.8%; Pred. No. 4.2e-17;
Matches 110; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 526 CTGCGATATTCGTAGTACGCGCTTTGACGCGACATTCACACTACGTTGGTGAATTC 585
||| |
Db 273 CTAAACAGTGTGATGAGCGCTTTGACGCGACATTCACACTACGTTGGTGAATTC 332
||| |
QY 586 TTCTGGGCGATTCGACGAGTCTGCTCCAGGCGATATTCAGAGTCTGCTACATCG 641
||| |
Db 333 TTCTGGGCGATTCGACGAGTCTGCTCCAGGCGATATTCAGAGTCTGCTACATCG 388

LOCUS 1977 bp mRNA PRI 19-DEC-1995
DEFINITION Human alpha 7 neuronal nicotinic acetylcholine receptor mRNA, complete cds.
ACCESSION U40583
NID 9125076
VERSION U40583.1 GI:1125076
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 1977)
AUTHORS Logel, J., Dreling, C., Barnhart, M., Antle, C. and Leonard, S.
TITLE Nucleotide Sequence and Transcript Size of the Alpha-7 Neuronal Nicotinic Acetylcholine Receptor in Human Postmortem Brain
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1977)
AUTHORS Leonard, S.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1995) Sherry Leonard, University of Colorado Health Sciences Center, C-268-71 Pharmacology, 4200 E. Ninth Ave, Denver, CO 80262, USA

FEATURES
source
1. 1977
Location/Qualifiers

1. 1977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="pBSalpha7"
/clone_lib="Clontech lambda gt10 cDNA library: human hippocampus"
/tissue_type="hippocampus"
/sex="male"
/dev_stage="20 yr old adult"
1. 73
8. 1516
/codon_start=1
/product="alpha 7 neuronal nicotinic acetylcholine receptor"
/protein_id="AA83561.1"
/db_xref="PIR:G1125077"
/db_xref="GI:1125077"
/translation="WRCSFGGVALAASLHVSIGEFQRLKYLKLVNPLERPV
ANDSOPLVFSLNLIQIMDVDEKNQVLTINIMLQMSWTDLQNVSEYPGVKTYR
PDGOIMKPDILLVNSADERDPAFTNVLVNSGHCQYLPQGIFFKSCYIDRWPFQ
VQHCRIKGSWYSGMSLDQMOEADISGVIENGENDIVIGIKRSERYECCKEPP
DYTFVTMRRLRYGLNLIPCVLSALVFLPADSGEKISLITVLSTVPM
LVVAIMPATSDVPLIAQFAS TMIVGLSVYVTVIVLQHHHDPDGMKPKMTVILLN
MCAMFLRKMRGEGEDKVRPAC OHKORSSSLASVSAVAPPASNGNLIVIGFGD
GVHCVPTPDGGVCGAMPAPT HDEHLHGQGPPEGDPLAKILEEVRVIANRFRCQ
DESAVQSEMKFACVVDRLCLMAFSVFTICTIGILMSAPNFVEAVSKDFA"
1514. 1977

3'UTR 1954 460 t
BASE COUNT 426 a 567 c 524 g
ORIGIN

